EXPOSURES TO MIXTURES OF AIR POLLUTANTS: ANALYSIS OF BIOLOGICAL, PERSONAL AND AREA MONITORING

by

Feng-Chiao Su

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Environmental Health Sciences)
in the University of Michigan
2013

Doctoral Committee:

Professor Stuart A. Batterman, Chair
Associate Professor Bhramar Mukherjee
Assistant Professor Sung Kyun Park
Professor Thomas G. Robins

DEDICATION

To my parents and brother,
Lien-Chen Su, Shui-Shu Chang and Sheng-Pei Su,
who made the completion of this work possible,
for their endless love and support.

以此論文謹致我的父母(蘇連成，張水束)與弟弟(蘇聖斐)
感謝他們一直以來的關愛與支持
使這一切得以成就

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to many people for their support and assistance in the preparation of this dissertation. First, I would like to deeply acknowledge my advisor, Dr. Stuart Batterman, for his strong support, patient guidance and mentorship, and good humor throughout my PhD training. I particularly thank him for his extreme patience to teach me and to revise my writing again and again. Second, I would like to thank my committee members, Drs. Bhramar Mukherjee, Sung Kyun Park, and Thomas Robins for their contributions and suggestions over the past years (Dr. Mukherjee especially, who provides me advice on many statistical analyses). Additionally, Dr. Chunrong Jia at the University of Memphis, who proposed the project included in my dissertation, has been very helpful in solving problems I encountered. Susan Crawford, who is the student administrative coordinator at the Department of Environmental Health Sciences, helps me with many administrative affairs.

In addition to "professional and academic support", I am also very grateful for the friendship I have in my life. I would like to show my appreciation for my labmates, Jo-Yu Chin, Kai Zhang, Christopher Godwin, LiuLiu Du, Lei Huang, Huda Elasaad, Sergei Chernyak, and Andrew Ekstrom, especially Jo-Yu, who provides me with assistance in my research and life, as well as Chris, who polishes up my dissertation. Also, I am so indebted to Shi Li for solving and helping many statistical problems. The section of mixture distributions in this dissertation is a part of his wonderful works. I am very thankful to the fellows in 5L2F, AACCC (Dry parents, Yi-An Ko and Katie Hsu in particular), and all my friends in Ann Arbor and Taiwan for their care and concern.

Most importantly, I am very blessed to have my family, who give me unconditional and endless support. I truly appreciate that my parents encourage me to pursue my doctorate degree (which is also my dream) even though they know that I have to stay far away from them. Also, I am deeply grateful to my brother, who took care of me and accompanied me in Ann Arbor through the hardest time in my life.

Finally, I appreciate the financial support from the Education Department of the Government of Taiwan (Study Abroad Scholarships), the University of Michigan Center for Occupational Health and Safety Engineering (Project: VOC Exposures among a Nationally Representative Sample of US Workers: Analysis of the NHANES 1988 through 2004 Data Sets), the Health Effects Institute (Project: Modeling and Analysis of Personal Exposure to Pollutant Mixtures: Further Analysis of the RIOPA Data), and the University of Michigan Rackham Graduate School (International Students Fellowship and One Term Dissertation Fellowship).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1,4-DCB	1,4-dichlorobenzene
A-D	Anderson-Darling
AER	air exchange rate
AIC	Akaike information criterion
AQS	Air Quality System
BDCM	bromodichloromethane
BIC	Bayesian information criterion
BTEX	benzene, toluene, ethylbenzene, xylenes
CDF	cumulative distribution function
CTC	carbon tetrachloride
DBCM	dibromochloromethane
DF	detection frequency
DPM	Dirichlet process mixture
EM	expectation maximization
EPA	Environmental Protection Agency
ETS	environmental tobacco smoke
EVT	extreme value theory
GEV	generalized extreme value
GM	geometric mean
GOF	goodness-of-fit
GSD	geometric standard deviation
IQR	inter-quartile range
K-S	Kolmogorov-Smirnov
K-W	Kruskal-Wallis
LMM	linear mixed-effect 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 tertiary-butyl ether
NATA	National-Scale Air Toxics Assessment
NEI	National Emissions Inventory
NHANES	National Health and Nutrition Examination Survey
NHAPS	National Human Activity Pattern Survey
PAF	population attributable fraction

PAH	polycyclic aromatic hydrocarbon
PAMS	Photochemical Assessment Monitoring Stations
PBDE	polybrominated diphenyl ether
PERC	tetrachloroethylene
PMF	positive matrix factorization
PSU	primary sampling unit
QR	quantile regression
R^2	reduction in residual variance
RfC	reference concentration
RIOPA	Relationships of Indoor, Outdoor, and Personal Air
SBC	Schwarz Bayesian Information Criterion
SD	standard deviation
SPME	solid-phase microextraction
TCE	trichloroethylene
THM	trihalomethane
TVOC	total volatile organic compound
UATMP	Urban Air Toxics Monitoring Program
URF	unit risk factor
VOC	volatile organic compound

ABSTRACT

Introduction. Emission sources of volatile organic compounds (VOCs) are numerous and widespread. Concentrations of VOCs indoors typically exceed outdoor levels, and most people spend nearly 90% of their time indoors. Thus, indoor exposures 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, e.g., asthma, liver and kidney dysfunction, neurological impairment, and cancer. Although exposures to most VOCs for most persons fall below health-based guidelines, a subset of individuals experience much higher exposures. Thus, exposure to VOCs remains an important environmental health concern.

Important gaps remain in our understanding of VOC exposures. Generally, concentration and especially exposure data are limited. Like much other environmental data, VOC exposure data can show multiple modes, heavy tails, and sometimes a large portion of data below method detection limits (MDLs). Field data also show considerable spatial or inter-individual variability, and information on long-term exposure trends is lacking. Additionally, typically exposure occurs as a mixture, and mixture components may jointly contribute to adverse effects. However, most pollutant regulations, guidelines and studies remained focused on single compounds, and thus may underestimate cumulative exposures and risks. Finally, while many factors are known to affect VOC exposures, many personal, environmental and socioeconomic determinants remain to be discovered.

To help answer these questions and overcome limitations of previous analyses, this dissertation utilizes several novel and powerful statistical techniques with analyses focused on two large datasets. The overall objective is to understand the nature and significance of exposures to VOCs by identifying and characterizing exposure distributions (including extreme values), exposure trends, exposures to mixtures (including dependencies), and exposure determinants.

Methods. VOC data were mainly drawn from two datasets: the Relationship between Indoor, Outdoor and Personal Air study (RIOPA), the National Health and Nutrition Examination Survey (NHANES). The RIOPA study collected outdoor, indoor and personal measurements in three U.S. cities from 1999 to 2001. Approximately 100 non-smoking households, adults and children in each city were sampled twice for 18 VOCs. More than 500 variables potentially associated with exposure were also collected. NHANES used a stratified, multistage, probability-based sampling design to collect nationally representative samples. Blood VOCs were measured for a subsample of adults for each cohort studied between 1988 and 2004, and personal VOC measurements were collected in 1999/2000.

To estimate extreme exposures, Gumbel and generalized extreme value (GEV) distributions were fitted to the top 5 and 10% of VOC exposures. Health risks were also estimated. Simulated extreme value datasets, following the fitted GEV, Gumbel and lognormal distributions for VOCs, were compared to observations. Mixture distributions using the traditional finite mixture of normal distributions and semi-parametric Dirichlet process mixture (DPM) of normal distributions were also fitted, and goodness-of-fit was evaluated using simulations.

VOC trends from 1988 through 2004 were evaluated using linear quantile regression (QR) models, which are more robust than ordinary linear models and can indicate changes at different quantiles. Linear QR models with adjustments for solvent-related occupations and cotinine levels were fitted to VOCs at the 50^{th}, 75^{th} and 95^{th} percentiles.

VOC mixtures in RIOPA were identified using positive matrix factorization (PMF) and by toxicological mode of action. Dependency structures of mixture components were examined using mixture fractions and copulas, which address correlations of multiple variables across their entire distributions, and evaluated using simulation. Cumulative cancer risks were calculated for mixtures, and results from copulas and multivariate lognormal models were compared to observations. The fractions of exposure attributable to the outdoor and home microenvironments were also estimated. Finally, exposure determinants were identified using stepwise regressions and linear mixed-effect models.

Results. Extreme value exposures typically were best fitted by 3-parameter GEV distributions, and sometimes by the 2-parameter Gumbel distributions. In contrast, lognormal

distributions significantly underestimated both the level and likelihood of extrema. Among the VOCs measured in RIOPA, 1,4-dichlorobenzene (1,4-DCB) posed the greatest risk of cancer, e.g., for the top 10% of exposures, the predicted lifetime excess cancer risk exceeded 10^{-4}, which represents an upper bound estimate of 100 cancer cases if one million people were exposed daily over their lifetime to the 90^{th} percentile 1,4-DCB concentration. NHANES had considerably higher concentrations of all VOCs with two exceptions (methyl tertiary-butyl ether (MTBE) and 1,4-DCB). Considering the full distribution models, the finite mixture of normals with two to four clusters, and DPM of normals had superior performance in comparison to the lognormal models. DPM distributions provided slightly better fit than the finite mixture of normals.

In NHANES, most VOCs showed decreasing trends at all quantiles, e.g., median exposures declined by 2.5 (m,p-xylene) to 6.4% (tetrachloroethene, PERC) per year over the 15 year period. Trends varied by VOC and quantile, and were grouped into three patterns: similar decreases at all quantiles (including benzene, toluene); most rapid decreases at upper quantiles (ethylbenzene, m,p-xylene, o-xylene, styrene, chloroform, PERC); and fastest declines at central quantiles (1,4-DCB). These patterns reflect changes in exposure sources, e.g., upper-percentile exposures may result mostly from occupational exposure, while lower percentile exposures arise from general environmental sources. Trends of VOC emissions and ambient concentrations are supportive of the exposure trends, although the data suggest the importance of indoor sources and personal activities.

Four VOC mixtures in RIOPA were identified by PMF, which represented gasoline vapor, vehicle exhaust, chlorinated solvents and disinfection by-products, and cleaning products and odorants. Typically, mixture fractions were heterogeneous, e.g., the compounds and fractions changed with the concentration of the mixture. Three mixtures were identified by toxicological mode of action, representing VOCs associated with hematopoietic, liver and renal tumors. Estimated lifetime cumulative cancer risks exceeded 10^{-3} for about 10% of RIOPA participants. This exceeds the range that is normally considered to be acceptable (from 10^{-6} to 10^{-4}). The dependency structures of the VOC mixtures fitted Gumbel and t copulas, both of which emphasize tail dependencies. The copulas reproduced both risk predictions and exposure fractions with a high degree of accuracy, and performed better than multivariate lognormal distributions.

The analysis of VOC determinants showed that exposures were affected by indoor concentrations, city, and some personal activities, household characteristics and meteorological factors. Home concentrations accounted for an average of 63 (MTBE) to 75% (carbon tetrachloride) of total exposure. For gasoline-related VOCs (e.g., benzene, MTBE), important determinants were city, attached garages, self-pumping of gas, wind speed, and house air exchange rate (AER). Odorant and cleaning-related VOCs (e.g., 1,4-DCB, chloroform) were associated with city, AER, house size and family members showering. Dry-cleaning and industry-related VOCs (e.g., PERC, trichloroethylene) were associated with city, residence water supply type, and visits to dry-cleaners. These and other relationships explained from 10 to 40% of the variation, and are consistent with known emission sources and the literature.

Conclusions. Exposure data feature extreme values, multiple modes, temporal changes, heterogeneous inter-pollutant dependency structures, and other complex characteristics. Advanced statistical methods can improve estimates exposures and risks, and are needed to develop control and management guidelines and policies. Both extreme value distributions and mixture models provided excellent fits to single VOC compounds (univariate distributions); copulas may be the method of choice for VOC mixtures (multivariate distributions), especially for the highest exposures, which poorly fitted with parametric models and may represent the greatest risk. Declining VOC exposures reflect the effectiveness of emission controls, while more rapid decreases in ambient concentrations suggests the importance of indoor sources, occupation, personal activities and other factors. The identification of exposure determinants, including the influence of certain activities and environments, provides information that can be used to manage and reduce exposures. These results extend our understanding of and ability to model VOC exposures.

CHAPTER 1

Introduction

This chapter presents background information for volatile organic compound (VOC) exposures and the objectives of this dissertation. Section 1.1 discusses the motivation for the research. Section 1.2 presents the findings, limitations, and unsolved issues in previous studies related to VOC exposures. Section 1.3 lists four specific aims in this dissertation to fill the research gaps. Section 1.4 shows the organization of this dissertation.

1.1 Motivation

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 Jr 2000). Important outdoor sources include industrial emissions and other *stationary sources*, vehicles and other *mobile sources*, gasoline service stations and dry cleaners considered as *area sources* (MDE 2010; Ling et al. 2011). Indoor sources include many building materials, cleaning products, cigarette smoke, adhesives, paint strippers, moth repellents, and water chlorination byproducts (Wallace et al. 1987; Wallace et al. 1989; ATSDR 1997a; Brown 2002; Singer et al. 2006; Weschler 2011; US EPA 2012b). In the U.S. and in many other countries, indoor concentrations of VOCs typically exceed outdoor levels (US EPA 2012b). Moreover, most people spend nearly 90% of their time indoors (US EPA 1989). For these two reasons, indoor exposures often constitute a large share, and often the dominant share, of VOC exposures for most individuals, at least for the non-occupationally exposed population.[1] Decreased smoking rates and restrictions on tobacco smoking, for example, may have lowered indoor concentrations and exposures of some VOCs more than changes in outdoor concentrations. Studies are needed to understand how outdoor and indoor sources contribute to personal

[1] The occupationally-exposed sector is not addressed in this dissertation. Workplace exposures to VOCs can be high in many occupations, e.g., mechanics, machinists, off-set printing press workers, painters, service station attendants, petro-chemical industry workers.

exposures of air pollutants, a major motivation of the Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study (Weisel et al. 2005a).

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 (Lippy and Turner 1991; Mendell 2007; Rumchev et al. 2007; Kim and Bernstein 2009; US EPA 2012a, b). Information regarding toxicity, drawn largely from occupational and animal studies, is available for a number of VOCs. Several elements of this dissertation use the RIOPA VOC measurements with dose-response information, specifically, the unit risk factor (URF, also called slope factor) for cancer risk, and the reference concentrations (RfC) for non-cancer endpoints.[2] For example, lifetime individual excess cancer risks are estimated by multiplying the lifetime (70 year) exposure by the URF specific to the VOC (US EPA 2009). The estimated risk was compared to *de minimis* or acceptable values, which typically range from 10^{-6} to 10^{-4}. Previous work based on the nationally representative 1999-2000 National Health and Nutrition Examination Survey (NHANES) has shown that exposures of most VOCs for most persons fall below current guidelines designed to be protective for both acute and chronic (cancer) effects (Jia et al. 2008). However, a subset of individuals experience much higher exposures that do exceed guidelines, e.g., the estimated lifetime cancer risk from benzene exceeded 10^{-4} for 10% of adults, and 16% of adults exceeded the same risk level for chloroform. Information on these high exposures is very limited. This topic is the focus of Sections 2.2.3 and 3.3 of this dissertation, which examines and model extreme values of VOC exposures.

1.2 Literature Review

Emissions and ambient concentrations of VOCs. In the U.S., emissions of many VOCs have declined in recent years, motivated by concerns regarding both the direct health effects of VOCs and their role in forming tropospheric ozone. Emissions have been lowered by substituting low emitting materials and processes, using controls such as catalytic converters, and shifting away from manufacturing jobs where solvent use was common. Based on the U.S. National Emissions Inventory (NEI), VOC emissions have been reduced by 35% from 1990 to 2005, or 2.3% per year, mainly due to controls on industry and on-road mobile sources

[2] This information is used to estimate risks in Sections 2.2.3.1 and 2.2.7.2, and to select mixtures for analyses in Section 2.2.6.2).

(US EPA 2010b). These and most other estimates of emission trends primarily use empirical and engineering factors, not actual measurements.

Decreased emissions have lowered ambient concentrations. A comprehensive review of air toxics data collected from 1990 to 2005 in the U.S. EPA's Air Quality System (AQS) showed that median levels of benzene, toluene, ethylbenzene, styrene, xylene and tetrachloroethyelene (PERC) declined by about 5 to 7% per year; chloroform by 1 to 4% per year; and 1,4-dichlorobenzene (1,4-DCB) by 0 to 9% per year, depending on the period (McCarthy et al. 2007). Benzene trends have also been examined by Fortin et al. (2005), who estimated an average decrease of 6.2% per year from 1993 to 2002 and 9.8% per year between 1994 and 1999, mainly using Photochemical Assessment Monitoring Stations (PAMS) data, and by U.S. EPA (2003a; 2007; 2010d), which showed decreases in urban areas of 8% per year from 1994 to 2000, 3% per year from 2000 to 2005, and 4% per year from 1994 to 2009. PAMS data are collected in the warmest portion of the year (the "ozone season"), and do not represent annual averages. Somewhat faster declines (9.8% per year) have been shown for quarterly averages of benzene in California from 1990 to 1995 (Hammond, 1998), and by data in the Urban Air Toxics Monitoring Program (UATMP), which has operated year-round since 1987, and which includes several sites located near busy roadways, commercial or industrial facilities (US EPA 2001). Ambient data are subject to variability from year-to-year changes in emissions, meteorology and sampling methodology, although long term declines across a number of periods are quite consistent and indicate the effectiveness of emission controls (McCarthy et al., 2007). However, ambient monitoring only partially explains exposure trends due to the little time most individuals spent outdoors and the strength of VOC sources in building and commuting environments.

VOC monitoring and exposure assessment. Personal measurements of pollutant concentrations, obtained using samplers carried by individuals, are generally believed to provide the data most relevant for exposure purposes. The RIOPA and NHANES datasets include such measurements. RIOPA also includes indoor (in participant homes) and outdoor (outside of these homes) measurements, and the VOC samples in RIOPA represent repeated measurements (sampled twice). Details on the data collected in RIOPA and NHANES are given in Section 2.1.

Exposures to pollutants can be estimated in many ways, but biomarker measurements often are considered the best exposure indicator since they account for multiple settings (e.g., indoor, outdoor and commuting environments), sources and exposure pathways (Ashley and Prah 1997). In urine, concentrations of VOCs strongly correlate to indoor levels (Wang et al. 2007). In blood, VOC concentrations have been associated with airborne levels, smoking and other activities, as well as individual characteristics such as gender and body mass index (Lin et al. 2008). Biomarkers have limitations, e.g., VOCs with rapid clearance (short biological half-lives) will reflect only recent exposures, thus observed relationships between airborne and biomarker concentrations depend on the variability of airborne levels, the duration of exposure and sampling periods, and clearance rates (Kwok and Atkinson 1995; Sexton et al. 2005; Lin et al. 2008). To date, quantitative and nationally representative trends using biomarkers have not been reported. Such analyses require the use of consistent methodologies, representative and large samples, and long study periods. NHANES, which has collected biological samples over several decades, can provide a good estimate of trends in VOC exposures for the U.S. population. This topic is the focus of Section 2.2.5.

VOC monitoring programs in the U.S. and elsewhere, including RIOPA and NHANES, measure only a subset of VOCs. Monitoring often focuses on 1-ring aromatic VOCs (e.g., benzene, toluene, xylene), smaller aliphatic compounds (n-hexane, heptane), and a few chlorinated compounds, e.g., trichloroethylene (TCE) and carbon tetrachloride (CTC). The RIOPA study, discussed below, includes several aromatic and chlorinated compounds, as well as d-limonene, α-pinene, β-pinene and methyl tert-butyl ether (MTBE). In general, little information is available regarding levels of and exposures to very volatile VOCs, more polar compounds, and lower volatility VOCs. This dissertation focuses only those VOCs measured in RIOPA and NHANES.

High exposures. As noted, the highest exposures may be most significant in terms of their potential to cause adverse health effects. The assumption of lognormality has been widely applied in the analysis of concentration and exposure data. However, lognormal distributions may inadequately characterize the highest observations in a dataset. For example, VOC distributions can have "heavy" right-hand tails, which clearly neither fit normal nor lognormal distributions (Su et al. 2012). In these cases, parametric models will underestimate the highest exposures and risks.

One approach to characterize such *extreme values* in a dataset uses *extreme value theory* (EVT), which describes the probability and magnitude of events with low probability and high consequence events (Lenox and Haimes 1996). A variety of EVT models have been developed, including the Gumbel distribution (Gumbel 1958), the Fréchet distribution (Fisher and Tippett 1928), and the Weibull distribution (Weibull 1951; Ang and Tang 1975). These three distributions, respectively called type I, II and III extreme value distributions, belong to the broad class of *generalized extreme value* (GEV) distributions, which use shape, location and scale parameters to fit the tails of a distribution (Jenkinson 1955). EVT distributions are univariate models (e.g., applying to one VOC) and not full distribution models (applying only to a tail of the distribution). Despite these limitations, EVT distributions have many applications, as described next.

EVT has been widely applied in engineering (McCormick 1981), finance (Embrechts et al. 1997), and hydrology (Katz et al. 2002; Engeland et al. 2004) and other fields. Some, but not many, environmental application have been published, e.g., estimating the likelihood of meteorological conditions (Hüsler 1983; Sneyer 1983), exceedances of thresholds relevant to dietary intake of pesticides and heavy metals (Tressou et al. 2004; Paulo et al. 2006), concentrations of metals Mn and Pb in blood (Batterman et al. 2011), deposition of pollutants in surface soils (Huang and Batterman 2003), and risks of leakage due to pipe corrosion (HSE 2002). Additional application for air pollutants include the exceedance of air quality standards (Surman et al. 1987; Hopke and Paatero 1994), exposures to ambient air pollutants (Kassomenos et al. 2010), indoor concentrations of radon (Tuia and Kanevski 2008), and VOC exposures in the NHANES subset mentioned earlier (Jia et al. 2008).

Sections 2.2.3 and 3.3 apply EVT theory to the VOC exposure data in the RIOPA dataset, and provide a critique of the approach.[3] The analysis of extreme values is further extended in Sections 2.2.7 and 3.7, which uses *copulas* to model *dependencies* among mixture components. This analysis also looks at tail behavior, the region of the distribution 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.

[3] Portions of this work have recently been published: Su FC, Jia C, Batterman S. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. *Atmospheric Environment* 62: 97-106.

Mixture distributions of VOC exposures. Environmental exposures of many VOCs (and other pollutants) at the population level, say across the U.S., can be viewed as *mixtures of distributions*.[4] A (typically small) fraction of the population experiences high concentrations due to specific exposure events, while a (typically large) fraction of the population encounters much lower concentrations (Jia et al. 2008; Batterman et al. 2011; Su et al. 2012). For the lower concentrations, often measurements fall below method detection limits (MDLs). These "*non-detects*," which represent 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 (Lubin et al. 2004; Antweiler and Taylor 2008). The statistical issues associated with analysis of data with MDL issues are well known (Taylor et al. 2001; Krishnamoorthy et al. 2009).

Due to the variation in source emissions, differences in the settings and environmental factors where exposures occur, and the measurement issues just noted, distributions of VOC concentrations can have multiple modes, heavy tails, and significant portions of data falling below the MDL that 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. While 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 (Su et al. 2012), 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 datasets that are not adequately fit by a single common distribution, provide a flexible and powerful approach of representing the distribution of a random variable (Titterington et al. 1985;

[4] Note that mixture distributions (the subject addressed her and in more detail in Sections 2.2.4 and 3.4) are to be distinguished from VOC mixtures (addressed in Sections 2.2.6, 2.2.7, 3.6 and 3.7): the former applies to the nature of the distribution for a particular VOC; the latter applies to a combination of VOCs collectively observed as an exposure or concentration in a specific environment (e.g., residence). Some further subtleties in the nomenclature can arise in cumulative risk assessment, which deals with the potential toxicity of chemical or environmental mixtures, i.e., essentially simultaneous exposures to multiple chemicals (discussed in Section 2.2.6).

McLachlan and Basford 1988; McLachlan and Peel 2000). As examples, the *finite mixture of normal distributions* applies a set of "mixing weights" to a specified and finite number of component distributions, while the nonparametric *Dirichlet process mixture* (DPM) *of normal distributions* relaxes the need to pre-specify the number of component distributions and is potentially advantageous in terms of handling smoothing, modality and uncertainty (Escobar 1994; Mueller and Quintana 2004). Mixture of normals distributions have been extensively used in a variety of important and practical situations, although environmental applications have been very limited (Burmaster and Wilson 2000; Razzaghi and Kodell 2000; Taylor et al. 2001; Chu et al. 2005). This is the subject of Sections 2.2.4 and 3.4 of this dissertation.

Exposure assessment to VOC mixtures. *Environmental mixtures* have been defined as the combination of two or more chemical components, regardless of the sources or the spatial or temporal proximity where exposures occur (US EPA 1986). Environmental exposures typically involve mixtures of pollutants that occur either simultaneously or sequentially, and over both short and long periods. While there is growing interest and concern regarding the cumulative effects of mixtures, most pollutant standards, regulations and guidelines historically and for the most part remain focused on single pollutants compounds rather than mixtures of pollutants. There are several notable exceptions. For example, environmental regulations control airborne exposures to particulate matter and diesel exhaust (US EPA 2012a, d); occupational exposure limits exist for gasoline vapor (as well as its several of its components, e.g., benzene) (ACGIH 2012); and drinking water regulations collectively limit the four trihalomethanes (THMs) (US EPA 2013).

As noted earlier, if mixture components 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 mixture exposures 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 use interaction or additive assumptions among the mixture components. Following the methods recommended to analyze cumulative risks of mixtures (US EPA 2000b, 2003; ATSDR 2004), mixture components can be considered to have independent toxicities, meaning that each chemicals has a different mode of action and that the overall response is obtained by adding responses of each component, which is called *response addition* (Bliss

1939). For example, cumulative risks of cancer have been estimated using response addition across 13 VOCs (e.g., benzene, 1,3-butadiene, chloroform, formaldehyde, styrene, acetaldehyde, etc), and 6 metals (chromium VI, nickel, arsenic, lead, cadmium, and beryllium) (Sax et al. 2006). If mixture components have similar toxicity effects or mechanisms, then doses can be added, called *dose addition*. An example of dose addition is the use of toxic equivalency factors for polycyclic aromatic hydrocarbons, which relate the relative potency of compounds in the mixture to a reference compound, e.g., benzo(a)pyrene, which are used as weights in summing doses or concentrations in an estimate of the mixture's toxicity (US EPA 1993). U.S. EPA (1986) suggests that if interaction information is unavailable, then the additive assumption should be adopted. Sections 2.2.7.2 and 3.7.3 in this dissertation use such methods.

The understanding and analysis of environmental mixtures can be aided by several additional definitions. Three classes of mixtures have been defined (ATSDR 2004): (1) generated mixtures composed of compounds which 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 or stored and reach the same target population, e.g., metals, solvents and semivolatile wastes at Superfund sites. Generated and intentional mixtures may be common in some settings, for example, in workplaces and homes. However, exposure to multiple air pollutants emitted from different outdoor sources, e.g., CO, $PM_{2.5}$ and benzene from vehicles, and SO_2 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 (< 10) of components. Often, such mixture 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 and copulas. The compositions of mixtures, including the relative concentrations of mixture components, can vary considerably. *Dependencies* among components of exposure mixtures refer to the statistical relationships among the concentrations of each component in the mixture, and potentially to the composition of the

mixture. In general, the most common indicator of dependencies between two variables uses correlation measures. These include Pearson correlation coefficients (r), which assume that variables are normally distributed (Rodgers and Nicewander 1988), and non-parametric correlation measures of dependence, most commonly rank correlation measures using Spearman's rho and Kendall's tau, which are robust with respect to outliers and can describe some non-linear relationships. As noted above, environmental exposures often are not normally distributed, but can contain extreme values and can remain right-skewed even after log-transformation (Jia et al. 2008). Thus, parametric correlation measures can have significant limitations. Both types of correlation measures show only pair-wise dependencies, e.g., not those involving three or more variables, and may not be reliable indicators in the presence of non-linear associations (Schmidt 2006; Staudt 2010).

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

While there have been few environmental applications, copulas 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 shown lacking (Cherubini et al. 2004; Jean-Frédéric et al. 2004). As noted earlier, given that environmental exposures also involve non-normal distributions and extreme values (Jia et al. 2008; Su et al. 2012), copulas could be a good tool to explore dependency structures of multivariate exposures. 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. The application of copulas to the RIOPA VOC dataset is addressed in Sections 2.2.7 and 3.7 of this dissertation.

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 affecting health at individual and community levels can be classified into three groups: social/economic environment, the physical environment, and a person's individual characteristics and behaviors (WHO 2012). In this dissertation, parallels are drawn from these definitions by considering *determinants of exposures*, that is, factors affecting concentrations and exposures. Like health 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 house age), and lastly into personal factors (e.g., race/ethnicity, and behavior). While not entirely exclusive, these groupings provide a structure that may help the understanding and analysis of factors affecting exposure.

VOC exposures can vary tremendously among individuals. This variation appears to be driven largely by house-to-house variability, as compared to seasonal, neighborhood or measurement variability (Jia et al. 2011). In addition to this interpersonal or *spatial variability*, *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 VOC emissions, e.g., emission controls and process changes on both stationary and mobile sources (US EPA 2010b), which partially explains the decline in VOC exposures (Su et al. 2011). Simultaneously, indoor VOC concentrations have fallen in many buildings, a result of reduced or eliminated tobacco smoking, low VOC paints, and other indoor air quality improvements. *Short-term variability* can include effects of weather, season, personal activities and other factors, and relevant time frames can range from perhaps seconds to days. While these general effects are known, the identification of the factors causing VOC exposures, that is, exposure determinants, remains unclear. This is the subject examined in Sections 2.2.9 and 3.9 in this dissertation using the RIOPA dataset, which collected a more complete set of potential determinants than most or possibly all other VOC studies.

A review of 12 studies that examined VOC determinants is summarized in Table 1. (This review emphasized general, i.e., non-occupationally-exposed, populations.) The number of determinants is large and includes many *environmental determinants*. Elevated exposures have been associated with low ventilation rates and closed windows (Sexton et al.

2007; D'Souza et al. 2009; Riederer et al. 2009; Symanski et al. 2009; Wang et al. 2009), house type (apartment and mobile homes have higher benzene and chloroform levels than single family houses) (Riederer et al. 2009; Byun et al. 2010), fewer years lived in home or newer houses (associated with higher BTEX exposure (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 supplies often experience higher chloroform exposure than households using well water (D'Souza et al. 2009). In Korea, children had higher exposure to traffic-related VOCs, e.g., toluene, ethylbenzene, and m,p-xylene in the city with narrower streets and mixed walkways and driveways that increased proximity to traffic (Byun et al. 2010).

A modest number of *personal determinants* have been identified. VOC exposure has been related to ethnicity, e.g., Hispanics had higher exposure to benzene, toluene, ethylbenzene, xylene (BTEX), MTBE, and 1,4-DCB, Blacks had higher exposure to 1,4-DCB, PERC and chloroform (Riederer et al. 2009; Wang et al. 2009), and Mexicans had higher exposure to benzene and 1,4-DCB (Wang et al. 2009). Occupation clearly affects exposure, e.g., BTEX exposure has been linked to service station and vehicle repair jobs (Jo and Song 2001), and pinene, limonene, toluene, ethylbenzene and styrene have been associated with cleaning jobs (Wolkoff et al. 1998). However, effects of occupation on VOC exposures for the general public have rarely been observed. Machine-related jobs have been linked to BTEX exposure (D'Souza et al. 2009), and time at work/school has been associated with benzene, ethylbenzene, xylene and PERC exposure (Wang et al. 2009).

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

Few *socioeconomic determinants* have been identified. Education and income has been negatively associated with exposures of benzene, 1,4-DCB, PERC and chloroform (Wang et al. 2009). This might suggest that persons of higher socioeconomic position experience fewer high-exposure activities, e.g., house cleaning, reside in cleaner homes and neighborhoods (e.g., distant from traffic), and/or commute and work in cleaner environments. In the NHANES VOC dataset, Hispanic and Black adults had higher levels of BTEX, MTBE and 1,4-DCB after controlling for a environmental and personal covariates, suggesting possible cultural differences (D'Souza et al. 2009). In broad terms, many socioeconomic factors are expected to be correlated with yet to be identified environmental factors, which may be considered more direct determinants of concentrations or exposures. Thus, the identification of socioeconomic determinants may lead to increased understanding of VOC exposures, and may raise factors and hypotheses that can help to explain exposures.

While many exposure determinants have been identified, the underlying studies have several limitations, the significance and applicability of the determinants are uncertain, and many determinants likely remain undiscovered. First, many of the studies used small samples, e.g., the Birmingham study enrolled only 12 adults (Kim et al. 2002), the New York City study had 46 high school students (P Kinney et al. 2002), and the 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 (NCHS 2012b), outdoor and indoor concentrations, time activity, and other information was not collected. However, as mentioned, the RIOPA (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.

1.3. Research Objectives

The overall objective of this dissertation is to understand the nature and significance of exposures to VOCs though identifying and characterizing exposure distributions, exposure trends, exposures to pollutant mixtures, inter-pollutant dependencies, and exposure determinants. As discussed in Section 1.1 and 1.2, this objective is motivated by gaps in our understanding of exposures and current needs in exposure science and risk assessment. The work provides new analyses of the RIOPA and NHANES datasets with the objectives. There are four main aims, each with specific hypotheses, as described below.

Aim 1 addresses the characterization of full and extreme value distributions, with the hypothesis that a combination of standard and extreme value distributions can best characterize the distribution of pollutant exposures. Work included fitting univariate full distributions for outdoor, indoor, and personal VOC observations, fitting extreme value distributions to the highest 5 and 10% of measurements for each VOC, and estimating risks of extreme value exposures. The results include a comparison of distributions fitting for the RIOPA and NHANES studies. Additionally, mixture distribution models were developed that represented full distributions -- ranging from the lowest to the highest exposures. These take into account values below detection limits, extreme values, and values in the middle of the distribution into account.

Aim 2 examines changes over time in VOC exposures, based on VOC measurements in blood from 1988 through 2004 among a nationally representative sample in NHANES. Long-term trends have rarely been examined. The hypothesis is that exposures of most VOCs have declined over the past two decades due to product substitution and better emission controls.

Aim 3 provides an analysis of exposure mixtures with the goal of increasing understanding of exposures to multiple pollutants, especially for highly exposed individuals. We hypothesize that copulas and other advanced techniques that represent multivariate exposure distributions can allow accurate and efficient modeling of mixtures, joint distributions and dependency structures. This task focuses on identifying common/priority mixtures of different pollutants and evaluating their effects and significance. Exposure mixtures were selected on the basis of emission sources and toxicity followed by estimating the joint distributions and dependency structures of the mixtures.

Aim 4 investigates exposure determinants of VOC exposures, with the goal of investigating effects of indoor sources (e.g., smoking, attached garages, use of moth repellents), time activity information (e.g., time spent in outdoors, traffic), socioeconomic, demographic, meteorological and other factors. The hypotheses here are that indoor levels, environmental factors and personal activities can significant affect personal exposures, and that new relationships will be revealed using the RIOPA dataset. Linear mixed-effect models (LMMs) were used to identify sources and determinants of repeatedly indoor, outdoor and personal measurements. While QR models were originally proposed, we believed that linear mixed-effect models are more effective in identifying exposure determinants given the repeated measurements available in the RIOPA study.

1.4. Organization of This Dissertation

This dissertation is organized into four chapters: Chapter 1 (this chapter) summarizes the literature, defines specific terms, and states objectives of this research and its significance. Chapter 2 describes the data sources and statistical methods applied for each research aim. Chapter 3 presents the results and discussion for the four aims. Chapter 4 integrates the main findings of each research objective, and discusses implications. It also lists recommendations for further research.

Much of this work presented in this dissertation has been published in peer-reviewed journals. Primarily related to Objective 1, extreme value analysis (see Sections 2.2.3 and 3.3) has been published in *Atmospheric Environment* in 2012 (Su FC, Jia C, Batterman S. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. *Atmospheric Environment* 62(0): 97-106). In Objective 2, an analysis of VOC

trends (see Sections 2.2.5 and 3.5) has been published in *Atmospheric Environment* in 2011 (Su FC, Mukherjee B, Batterman S. 2011. Trends of VOC exposures among a nationally representative sample: Analysis of the NHANES 1988 through 2004 data sets. *Atmospheric Environment* 45(28): 4858-4867).

The rest of this work in the dissertation has been submitted to peer-reviewed journals. In Objective 1, an analysis of mixture distributions (see Sections 2.2.4 and 3.4) has been submitted to *Atmospheric Environment* in November 2012 (Li S, Batterman S, Su FC, Mukherjee B. 2013. Addressing extrema and censoring in pollutant and exposure data using mixture of normal distributions. *Atmospheric Environment*). In Objective 3, an analysis of VOC mixtures (see Sections 2.2.6, 2.2.7, 3.6 and 3.7) has been submitted to *Environment International* in February 2013 (Su FC, Mukherjee B, Batterman S. 2013. Modeling and analysis of personal exposures to VOC mixtures using copulas. *Environment International*). In Objective 4, an analysis of VOC determinants (see Sections 2.2.8, 2.2.9, 3.8 and 3.9) has been submitted to *Environmental Research* in February 2013 (Su FC, Mukherjee B, Batterman S. 2013. Determinants of personal, indoor and outdoor VOC concentrations: An analysis of the RIOPA data. *Environmental Research*).

CHAPTER 2

Material and Methods

This chapter describes the materials and methods used in this research. Section 2.1 introduces the two main datasets used, as well as several others. Section 2.2 describes the statistical approaches in the order of the four specific objectives (see Section 1.3).

2.1 Data Sources

2.1.1 Relationship between Indoor, Outdoor and Personal Air study

The RIOPA study contrasted three cities (Elizabeth, NJ; Houston, TX; Los Angeles, CA) that were expected to have different contributions from mobile and industrial emissions (Weisel et al. 2005b). Approximately 100 non-smoking households and non-smoking adults and children living in households in each city were recruited and studied from summer 1999 to spring 2001. Each of the household and participants was sampled twice about three months apart. Outdoor, indoor and personal air samples were collected using 48-hr 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, toluene, ethylbenzene, m,p-xylene, o-xylene, MTBE, styrene, 1,4-DCB, methylene chloride (MC), TCE, PERC, chloroform, CTC, d-limonene, α-pinene, β-pinene, 1,3-butadiene and chloroprene). Data for 1,3-butadiene and chloroprene were not reported due to low recovery. We excluded the MC measurements due to measurement issues (inconsistent blank contributions) (Weisel et al. 2005b). Styrene has higher uncertainty due to biased inter-laboratory consistency (Weisel et al. 2005c). A new variable, TVOC (total volatile organic compounds), was defined as the sum of the remaining 15 VOCs. MDLs ranged from 0.21 (α-pinene and PERC) to 7.1 (toluene) μg m^{-3}, and detection frequencies for the outdoor measurements ranged from 6.3 (β-pinene) to 96.8% (CTC), for indoor measurements ranged from 25.8 (TCE) to 95.5% (CTC), and personal measurements ranged from 22.5 (TCE) to 96.7% (CTC) (Weisel et al. 2005b). Measurements below the MDLs were replaced with

one-half of this value. Further details of RIOPA and its design are provided elsewhere (Weisel et al. 2005a).

RIOPA participants were administered three questionnaires, from over 500 variables were derived. A baseline questionnaire addressed demographics 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, household air exchange rates (AERs), type of building, and existence of attached garage); and a third questionnaire collected time activity information, e.g., time spent indoors at school/work, pumping gas, bathing or showering, and gardening (Weisel et al. 2005a). Geographic and meteorological information (e.g., city, outdoor temperature, wind speed, and relative humidity) was also obtained for each household.

2.1.2 National Health and Nutrition Examination Survey

For biological VOC samples, data were obtained from two cohorts of NHANES III (1988-1991, 1991-1994), and three cohorts of "continuous NHANES" (1999/2000, 2001/2002 and 2003/2004). Initially, NHANES focused on health and nutrition issues and did not include contaminant measurements. Participants were selected to be nationally representative using a stratified, multistage, probability-based sampling design, e.g., elderly and minorities were over-sampled. VOCs were measured for a subsample of adults aged 20-59 years for each cohort studied between 1988 and 2004, with sample sizes from 605 to 1489 as shown in Appendices A and B, (NCHS 2000, 2010d). To obtain nationally representative results and allow comparability between cohorts, each cohort used the same sampling and weighting scheme (NCHS 2006). There are several differences between cohorts. NHANES III used a 6 year survey cycle, 81 primary sampling units (PSUs) from 1988 to 1994 (randomly divided into two groups for 1988-1991 and 1991-1994), and about 15,000 participants per cohort. Continuous NHANES used a 2 year survey cycle, 12 PSUs in 1999/2000 (3 PSUs were omitted due to delays in data collection), 15 PSUs in both 2001/2002 and 2003/2004 cycles, and approximately 10,000 participants per cohort (NCHS 2010a, 2010b, 2010c). Thus, continuous NHANES encompassed fewer PSUs and obtained smaller samples, and consequently, standard errors may be larger than those in NHANES III (NCHS 2006).

NHANES III and continuous NHANES used similar procedures to collect and analyze blood samples (NCHS 2000, 2011). Participants arrived at a central location and designated time, and were then shepherded through four air conditioned trailers that comprised the mobile examination center (MEC) in visits that could require up to 4 hr (NCHS 2009). Blood samples were drawn in the third trailer. Whole blood samples were analyzed for 15 compounds: benzene, toluene, ethylbenzene, m,p-xylene, o-xylene, styrene, chloroform, bromodichloromethane (BDCM), dibromochloromethane (DBCM), bromoform, 1,4-DCB, PERC, MTBE, CTC, and TCE. Analyses used purge-and-trap extraction or headspace solid-phase microextraction (SPME), and capillary gas chromatography/mass spectrometry. Consistent quality control and quality assurance protocols were maintained (NCHS 2010e).

For airborne personal VOCs, the 1999/2000 cohort of the NHANES, which included personal VOC measurements for 851 participants (NCHS 2012a), was used to compare with the RIOPA study. The RIOPA and NHANES studies shared ten VOCs in common (benzene, toluene, ethylbenzene, m,p-xylene, o-xylene, MTBE, 1,4-DCB, TCE, PERC and chloroform). While the recruitment strategy and study purposes differed, NHANES and RIOPA used similar sampling methods and periods (48 to 72 hr for NHANES) as well as study periods. In NHANES, four observations were deleted (two cases, participant ID = 468 and 578, that had excessively long sampling periods, and two cases, participant ID = 3852 and 4076, with extremely high concentrations of benzene, xylenes or toluene), also described by Jia et al. (Jia et al. 2008).

2.1.3 Other Datasets

Several datasets were reviewed to derive trends in nationwide emissions and ambient concentrations to compare to the NHANES measurements. Emission data were taken from the National-Scale Air Toxics Assessment (NATA), an ongoing program used to derive pollutant emissions and risks (US EPA 1996, 1999a, 2002). Trend analyses using emission inventory must account for changes in inventory methods, e.g., NATA included additional source types in 1999 (US EPA 1999a). We also used NATA's dispersion model predictions for 1996, 1999 and 2002, which are based on the NATA emission data but which reflect effects of dispersion. NATA significantly underpredicts concentrations of many VOCs, due to missing and underestimated emission sources, among other reasons (US EPA 2010a). However, our

analysis stressed relative changes, which may be less sensitive to these biases. Several ambient monitoring datasets were also reviewed, including the 1993 to 2004 aromatic concentrations in the PAMS (US EPA 2011a), the 2001 to 2004 data from UATMP (US EPA 2001), and the 1990 to 2004 data from AQS (US EPA 2011a). PAMS and AQS data cover or nearly cover the period spanned by the five NHANES cohorts. Site annual averages from the AQS were downloaded and national level annual averages were calculated. To obtain reliable and representative averages, only sites collecting 24-hr samples were used, each site had to collect at least 24 measurements per year, and at least 20 sites meeting these criteria were required to compute the annual average. Trends were plotted and percent changes per year were calculated using simple linear regressions.

2.2 Statistical Methods and Data Analysis

2.2.1 Descriptive Analyses

The detection frequency (DF), defined as the percentage of measurements exceeding the MDLs, excluding missing values, was calculated for each VOC in both RIOPA and NHANES datasets (see Supplemental Table S1 and S2).

2.2.1.1 RIOPA Data

Descriptive statistics were calculated for all VOCs, including sample size, mean, standard deviation (SD), geometric mean (GM), geometric standard deviation (GSD), minimum, 25^{th}, 50^{th}, 75^{th}, 95^{th} percentiles, and maximum; these were calculated for all measurements (outdoor, indoor, and personal), and also stratified by city. Spearman rank correlations were also calculated for the VOC variables.

2.2.1.2 NHANES Data

Descriptive analyses followed the NHANES analytic guidelines (NCHS 2006) and used weights to account for NHANES' hierarchical clustered sampling strategy. VOCs with very low (<5%) DFs across the five cohorts were excluded from further analyses. MTBE was only measured in continuous NHANES, and was excluded from certain analyses. To ensure a sufficient sample size, at least 300 observations per VOC per cohort were generally required. New variables formed to examine related groups of VOCs included BTEX (the sum of benzene, toluene, ethylbenzene, m,p- and o-xylene concentrations) and total THMs (ΣTHM,

the sum of chloroform, BDCM, DBCM and bromoform). Spearman rank correlation coefficients were used to test associations among blood VOCs and among the air and blood measurements for the 1999/2000 cohort. Group differences in key demographic variables (age, gender, race, education levels, and income) among the cohorts were tested using ANOVA and Chi-square tests for continuous and categorical variables, respectively.

2.2.2 Full Distribution Fitting

Maximum likelihood estimates (MLEs) were used to fit the full distribution of each VOC, and goodness-of-fit (GOF) was examined using Anderson-Darling (A-D) tests (Haas 1997) with the following candidate distributions: beta general, chi-square, Erlang, exponential, extreme value, gamma, inverse Gaussian, logistic, log logistic, lognormal, normal, Pareto, Pearson type 5, Rayleigh, Student, triangular, uniform, and Weibull. The null hypothesis for the A-D test is that VOC observations come from a specific distribution. The A-D test, a modification of the Kolmogorov-Smirnov (K-S) test, emphasizes tail behavior (Stephens 1974), so it is more appropriate for evaluating environmental exposure data which are usually right-skewed distributions. Graphical examinations also provided insight. For each VOC measurement type (outdoor, indoor, adult personal, child personal), all observations (i.e. both first and second visit samples) before and after log transformation were used for full distribution fitting.

Full distribution fitting for VOC observations were performed using @Risk and the Decision Tools for Excel (Palisade Corporation, Ithaca, NY).

2.2.3 Extreme Value Analyses

2.2.3.1 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 US EPA Integrated Risk Information System (US EPA 2012a), the Office of Environmental Health Hazard Assessment's 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 Table 2, along with the reference concentration (RfC) and toxic endpoints. URFs are not available for toluene, m,p-xylene, o-xylene, d-limonene, α-pinene and β-pinene. The two visit measurements for each adult in

RIOPA were averaged as an estimate of the long-term exposure concentration. The excess individual lifetime cancer risk for a specific VOC i was calculated as:

$$R_i = C_i \ URF_i \quad (1)$$

where R_i = excess individual lifetime cancer risk (probability), C_i = concentration ($\mu g\ m^{-3}$), and URF_i= unit risk factor (cancer cases per $\mu g\ m^{-3}$).

Following guidance for mixtures (US EPA 2000a), risks were calculated by response addition for those VOCs that cause the same toxic effect on same target organ. In this case, results of eq. (1) were summed for each participant for the several chemicals in the mixture. Three mixtures were considered: 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 TVOC (Borgert et al. 2004; IARC 2012). TVOC also serves as a general indicator of VOC exposure, and can be used to identify the dominant contributors to VOC risks. The cumulative risk of mixture exposure was computed for each subject by summing the risks of components in the mixture, and extreme values of the cumulative risk were taken as the top 5% and top 10% of this sum over all persons.

2.2.3.2 Gumbel Distribution Fitting

Gumbel distributions were first used to estimate extreme value distributions for the top 5 and 10% of all observations and all measurement types. The sample size for the child personal samples was smaller (n=209) than the other measurement types (indoor, outdoor and adult-person measurements had a typical n=550), thus only the top 10% of the observations were considered as extrema for child personal exposures. A probability plot method was used to fit the Gumbel distributions as follows (Barnett 1975). First, extrema were ranked in descending order. Then, each observation was plotted against -ln[-ln(Pv)], where Pv was computed as:

$$Pv = (r - 0.44)/(n + 0.12) \quad (2)$$

where r = reverse rank of VOC concentrations, and n = sample size. This method allows GOF to be visualized as agreement to a regression line, and quantitative agreement is noted by the regression's R^2 statistic.

2.2.3.3 Generalized Extreme Value Distribution Fitting

To focus on health risk of the highest VOC exposures, further extreme value analyses were applied to personal VOC observations in RIOPA and the results were compared with the NHANES dataset. A broader class of extreme value distribution, the GEV distribution (Jenkinson 1955), was fitted to each extrema dataset (5 and 10% cut-offs for VOC exposures). The GEV probability density function is expressed as:

$$f_{\xi,\mu,\sigma}(x) = (((1+(\xi(x-\mu)/\sigma))^{-1-1/\xi})/\sigma)\ \exp(-(1+(\xi(x-\mu)/\sigma))^{-1/\xi}) \qquad \text{if } \xi \neq 0 \tag{3}$$

where ξ = shape parameter, μ = location parameter, σ = scale parameter, and x = data observation. If $\xi > 0$, the GEV distribution belongs to Fréchet family; if $\xi < 0$, the GEV distribution belongs to Weibull family (Jenkinson 1955); and if $\xi = 0$, the GEV distribution belongs to the Gumbel family, which permits simplification of eq. (3):

$$f_{0,\mu,\sigma}(x) = ((e^{-(x-\mu)/\sigma})/\sigma)\ \exp(-e^{-(x-\mu)/\sigma}) \tag{4}$$

The three parameters of the GEV distribution were determined by MLE, and GOF was examined using A-D tests with the null hypothesis that data subset comes from the GEV distribution. The A-D test, a modification of the K-S test, emphasizes tail behavior (Stephens 1974), so it is the most appropriate for evaluating extreme value distributions. Empirical A-D test p-values were calculated for the repeated (bootstrap) samples in the NHANES weighted dataset.

For GEV distribution fitting, only adult personal measurements were estimated because they should be the most representative of exposure. We selected adult subjects due to the larger sample size, namely, 544 measurements for 305 participants (299 and 245 measurements in first and second visits, respectively, of which 239 adults had valid samples in 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 risk of the long-term exposure was the most concerned (concentrations were too low for acute effects), the averaged measurements over the two visits were used. We next identified outliers, which initially were defined as a value twice that of the next highest observation, and also influential observations, identified as observations which clearly altered statistical results. Observations identified as

being both outliers and influential were excluded in subsequent analyses; these very few observations are noted. The sample sizes of the final top 5% and the top 10% of observed concentrations were 12 and 24, respectively.

2.2.3.4 Extreme Value Simulation

For further evaluation, simulated extreme value datasets (n=10,000) were generated for each personal adult VOC that followed the fitted GEV, Gumbel and lognormal distributions. Because lognormal distributions are commonly employed for exposure data, these distributions were fit to the full datasets by MLE, and the evaluation focused on extrema, again defined as the top 5% and top 10% of the full distribution. Simulated datasets were generated for the GEV and Gumbel distributions that matched the top 5% and top 10% of observations. Simulated data (n=10,000) were also generated for the lognormal distributions that matched the full distribution of observations. The simulated data were then compared to observations using K-S tests and graphical analyses, and p-values were estimated. Finally, in a risk assessment-oriented application, we compared the fraction of persons with cancer risks exceeding 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2} cut-offs for the three sets of distributions to observed fractions. These analyses were conducted for both individual VOCs and mixtures.

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

2.2.4 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 samples for adults were selected, primarily because the sample size for the adult cohort (n = 544 for each VOC) was largest, and because the personal samples should best reflect exposure. The two laboratories used to analyze samples had different MDLs. Since the use of two laboratories is somewhat unusual, all data under MDLs were replaced with a single value using 0.5 × the higher MDL. Because the VOC data in RIOPA had many extreme values (Su et al. 2012), the density estimation methods were implemented using logarithms of the concentration value, as described next.

2.2.4.1 Finite Mixture of Normal Distributions

Finite mixture distributions are commonly used to identify and model sub-populations within an overall population. Rather than identifying the sub-population that an individual observation belongs to, these models assume that the observed 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

$$f(y_i) = \sum_{k=1}^{K} \lambda_k f_k(y_i) \quad (5)$$

where f_k is the component density of the k-th 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

$$f(y_i) = \sum_{k=1}^{K} \lambda_k N(\mu_k, {\sigma_k}^2) \quad (6)$$

The finite mixture of normals represented by (6) is a potential choice for handling concentration and exposure data that can have multiple modes and heavy tails. Such normal mixtures are popular choices with attractive properties (Titterington et al. 1985). Since the mixtures are constructed as a linear combination 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 (Redner and Walker 1984; Titterington et al. 1985; McLachlan and Peel 2000). Since closed forms of MLEs of (5) are not available, mixture distributions are commonly fitted using expectation maximization (EM) type algorithms (Dempster et al. 1977; Meng and Pedlow 1992; McLachlan and Krishnan 1997). We used the EM algorithm and considered a constrained maximum likelihood method to estimate (6) with a further constraint that the location of the first cluster (generally the lowest) is under 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 selection of 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). While this criterion generally favors larger K, there is considerable practical support for its use due to simplicity (Fraley and Raftery 1998). The Bayesian information criterion (BIC) appears attractive due to their statistical properties as well as the simplicity of 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 the clusters' separation (Biernacki et al. 2000). In general, with limited amount of data, a corrected version of AIC such as AICc (Hurvich and Tsai 1989) may be preferable. For these finite mixture distributions, we fitted model (6) with K=2 to 5 clusters, and selected the optimal model based on AICc. This analysis was conducted for each of the three VOCs.

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

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

2.2.4.2 Dirichlet Process Mixture of Normal Distributions

Bayesian density estimation methods using Dirichlet process mixture (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 uncertainty about the density estimate (Ferguson 1983; Escobar 1994; Mueller and Quintana 2004). Instead of pre-specifying 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 of normal distributions assumes that these normal parameters θ_i follow a random distribution G generated from Dirichlet process (Ferguson 1973), which can be represented as:

$$\theta_i \mid G \backsim G \text{ i.i.d. and } G \mid \alpha, G_0 \backsim DP(\alpha G_0) \quad (7)$$

$\mathrm{DP}(\alpha\, \mathrm{G}_0)$ 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 provided the following representation for the leave-one-out conditional distributions (Blackwell and MacQueen 1973):

$$\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, \backsim \frac{\alpha}{n-1+\alpha} G_0 + \frac{1}{n-1+\alpha} \sum_{j \neq i}^{n} I_{\theta_j}(\cdot) \quad (8)$$

In this approach, $\theta = (\theta_1, \ldots, \theta_n)$ will be reduced to certain K distinct values ($K < n$) with positive probability. From (8), two well-known extreme cases of the DPM can be derived. As $\alpha \to \infty$, the DPM reduces to a parametric model, namely $\theta_i \backsim G_0$ independent and identically distributed (n clusters), whereas $\alpha \to 0$ implies a common parametric model, namely $\theta_1 = \cdots = \theta_n = \theta^*$ with $\theta^* \backsim G_0$ (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 of normals does not require specification of the number of clusters as needed for parametric mixture distributions, such as the finite mixture of normals discussed previously. In practice, suitable values of K will typically be small relative to the sample size n. The implicit prior distribution on K is stochastically increasing 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 choice of the Gamma parameters. Given the sample size in the test dataset (n=544), for prior information, $\alpha \sim \mathrm{Gamma}(0.3, 0.4)$ favors K=1-3 clusters; $\alpha \sim \mathrm{Gamma}(1.2, 2.5)$ favors 1-5 clusters; $\alpha \sim \mathrm{Gamma}(2, 1.5)$ favors 2-10 clusters; and $\alpha \sim \mathrm{Gamma}(5, 2)$ favors 5-20 clusters. A sensitivity analysis was conducted on these prior specifications.

Computational methods were followed that allowed the evaluation of posterior distributions for all model parameters and the number of components, and also 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.

2.2.4.3 Goodness of fit Criteria

Goodness of fit for the density estimation methods was determined by comparing the estimated cumulative distribution function (CDF) $\hat{F}_{est}$ to the empirical CDF $\hat{F}_{emp}$ based on the observed data. Although all observed/generated data were used to estimate the CDF by each method, goodness of fit was evaluated using only the data above the MDL. Both the mean squared error (MSE $= \sum_{i,y_i>MDL} [\hat{F}_{emp}(y_i) - \hat{F}_{est}(y_i)]^2 / \sum_i I(y_i > MDL)$), and the mean absolute error MAE $= \sum_{i,y_i>MDL} |\hat{F}_{emp}(y_i) - \hat{F}_{est}(y_i)| / \sum_i I(y_i > MDL)$ were considered. The estimated proportion of observations above the MDL, which is often termed the detection frequency, for empirical and estimated distributions was compared.

2.2.4.4 Simulation Study

For further evaluation of the mixture distributions, several forms of underlying true distributions and varying amounts of left censored data (below MDL) were considered as true generation models. Three methods were compared: a single normal distribution; a finite mixture of normals; and DPM of normals. Two underlying distributions with features similar to the three VOC samples from the RIOPA study were selected: a $\text{normal}(0, 2^2)$ and a mixture specified as $1/2\ \text{Gamma}(3, 1.5) + 1/2\ \text{Uniform}(-3, 8)$. The former is symmetric and the latter is right-skewed with heavy tails, and both have multiple modes when data under 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 described above) were calculated for each method, target distribution, and choice of P_0. A dataset of size n=1000 was generated for each simulation under each setting. The average values of MSE and MAE across 500 simulations are reported.

For the finite mixture of normals, 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%), possibly 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. Additionally, the MLE method for finite mixture models is susceptible to other problems, e.g., nonunique solutions (Redner and Walker 1984; Titterington et al. 1985; McLachlan and Peel 2000). Thus, data below the MDL was replaced by uniformly generated pseudo-data from U(0, MDL) if the finite mixture of normals did not converge. In contrast, all of the single normal and DPM method simulations converged.

2.2.5 Trend Analyses of VOC Exposures

Concentration trends were examined using quantile regression (QR) models, which estimate changes in conditional quantiles of a response variable with changes in VOC levels (Koenker and Bassett 1978). This semiparametric method makes no parametric distribution assumptions for random errors. Model coefficients are estimated by optimizing an objective function and the accompanying standard errors are derived using either parametric assumptions on the model coefficients or via resampling techniques, e.g., bootstrap analysis (Cade and Noon 2003). Compared to ordinary regression models, QR models are more robust, e.g., resistant to effects of outliers, a special concern for skewed distributions, which have been observed even after log-transformation of VOC data, following the NHANES guidelines (Jia et al. 2008; NCHS 2010g). Moreover, QR models indicate changes at different quantiles, e.g., allowing comparison of trends at median and upper percentiles, and exploration of exposure patterns. Linear QR models were fitted for 0.5, 0.75 and 0.95 quantiles (50^{th}, 75th and 95^{th} percentile concentrations). In a sensitivity analysis to allow changes in trend over the long interval (1994-1999) between the NHANES III and continuous NHANES cohorts, piecewise QR models were used with knots (locations where the slope changes) at several locations (e.g., 1991-1994, 1999/2000).

To facilitate interpretation, annual average percentage changes in untransformed (raw) concentrations were computed for each VOC and quantile, e.g., the change across the 15 year study period is $1/15\ (C_5 - C_1)/C_1\ 100\%$, where C_1 and C_5 are concentrations for a specific VOC and quantile in the first and fifth cohorts, respectively. Annual relative changes were calculated similarly for emissions and ambient concentrations.

Cigarette smoking is an important source of benzene and other aromatic compounds (L Wallace et al. 1987), and cotinine is a reliable biomarker of tobacco smoke (Benowitz 1999).

Correlations between serum cotinine levels and blood VOCs were determined, and the QR models were adjusted for this parameter.

Results of trend analyses might be affected by shifts in the occupational mix, e.g., the declining number of workers in industries where solvent use may be common. To account for such effects, we identified occupational groups associated with VOC concentrations, and adjusted QR models using indicator variables for these groups. Because many of the 41 occupational groups in NHANES had small sample numbers, groups were consolidated into eight categories (managerial and professional specialty occupations; professional specialty occupations; technical, sales and administrative support occupations; service occupations; farming, forestry and fishing occupations; precision production, craft, and repair occupations; operators, fabricators, and laborers; military occupations) based on 1990 Census Industrial & Occupational Classification Codes. Due to the small number of military personnel (n = 7), this category was dropped. ANOVAs were used to test whether VOC levels were associated with these occupational categories, using the managerial and professional specialty category as a reference group.

While the QR models used cohort-specific weights to obtain population-weighted results, these models cannot account for NHANES' cluster sampling. As a sensitivity analysis to evaluate the effect of clustering, trends in the mean were estimated using linear and piecewise models with the appropriate weights, and compared to regression results with and without adjustments for strata and clusters.

SAS 9.2 (SAS Institute, Cary, North Carolina, USA) was used for statistical testing and model development. Weighted analyses used Surveymeans and Surveyreg, and QRs used Quantreg. Other analyses were calculated using Excel (Microsoft, Redmond, WA).

2.2.6 Identification of Mixtures

2.2.6.1 Positive Matrix Factorization Analyses

VOC mixtures in the RIOPA dataset were selected using two approaches. The first approach identified common VOC mixtures using positive matrix factorization (PMF), a multivariate analysis that is similar to factor analysis, but with the ability to incorporate uncertainties on each measurement that potentially reflect sampling errors and MDLs (Paatero

and Tapper 1994; Anderson et al. 2001). Based on the uncertainty, variables are modeled as weak or strong, i.e., 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 TVOC were designated as "weak" given its higher uncertainty. Measurements below MDLs were retained, but assigned large uncertainties to reduce their influence (US EPA 2008a).

PMF decomposes two matrices from the sample data: a matrix of factor profiles, which represent the mass and percentage of each species apportioned to the factor, and a matrix of factor relative contributions, which gives the contribution of each factor to the total concentration of each observation (US EPA 2008a). Because there is no optimal or *a prior* manner for selecting the number of factors, multiple PMF analyses were conducted using with 3, 4 and 5 factors. Each was tested using GOF indicators, specifically, scaled residuals and Q values. The latter is the sum of squares of the residuals divided by the uncertainties for the concentrations of individual compounds (Anderson et al. 2001; US EPA 2008a).

To address seasonal variation, non-averaged VOC observations were grouped into warm (April to September) and cold (October to March) seasons, and PMF analyses were run separately for all groups. PMF analyses were run in various groups, and the final group (presented in this dissertation) separated indoor VOCs; outdoor VOCs, and combined adult and child personal VOCs. The logic for this arrangement was that different emission sources would dominate indoor, outdoor and personal measurements, although the same source types would affect personal measurements of adults and children, but in different amounts. Combining child and adult groups also increased sample size. Apportionments for adults and children could be separated after the analysis in order to resolve differences, e.g., children would not be expected to have occupational exposures. In addition, to avoid potential biases involved in repeated measurements (i.e., cluster effects) in further analysis (e.g., copula analysis), PMF analysis applied to the personal adult measurements collected at the first visit. 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 (US EPA 2008a).

To help understand the personal, behavioral and environmental variables associated with high exposure mixtures, a limited analysis using bivariate logistic regression models was undertaken. VOC mixtures identified using PMF were divided into high and low groups, using a cutoff of the 75th percentile of the mixture's total concentration (sum of each component). Candidate variables for the logistic regressions, based on earlier work that identified determinants of VOC exposure (Su et al. 2013), included city, ethnicity, employment status, the presence of attached garage, self-service pumping gas, open doors or windows, other family members taking showers, the use of fresheners, and household AERs. The logistic regression models used proc logistic in SAS 9.2 (SAS Institute, Cary, North Carolina, USA).

2.2.6.2 Toxicological Mode of Action

The second approach for selecting exposure mixtures used the toxicological mode of action, which considers the biochemical pathways and outcomes that may be affected by pollutant exposure (Borgert et al. 2004). Two mixtures were considered that had common cancer endpoints: (1) VOCs associated with hematopoietic cancers (lymphomas and leukemia), which include benzene, MTBE, 1,4-DCB, TCE and PERC; and (2) VOCs associated with liver and renal tumors, which include ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform and CTC (Borgert et al. 2004; IARC 2011). The two mode of action mixtures contained 5 and 7 components, respectively. It should be noted that selecting a mixture based on mode of action is a completely different approach from those determined using PMF or other correlation-based measures, which are driven exclusively by the pattern of occurrence.

To reduce the number and complexity of analyses in mixtures containing a larger 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 trimmed to a group of gasoline-related compounds (ethylbenzene and MTBE), and chlorinated hydrocarbons (1,4-DCB, TCE, PERC, chloroform and CTC). The analysis then proceeded with these groups.

2.2.7 Dependency Structures of Mixtures

2.2.7.1 Copula Analysis

Dependency structures of the identified mixtures (using personal adult measurements at the first visits) were fitted to copulas using MLEs and five candidate copulas (Gaussian, t, Gumbel, Clayton, and Frank). GOF tests were conducted using Akaike and Bayesian information criterion, 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:

$$C(u_1, u_2, \ldots, u_p) = \text{Prob}(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_p \leq u_p) \quad (9)$$

where C is a copula function, U_i, i=1,..p are uniformly transformed random variables corresponding to the marginal distribution functions $F_i(x_i)$, and p is the number of variables. The joint distribution function can also be expressed as:

$$C[F_1(x_1), F_2(x_2), \ldots, F_p(x_p)] = F(x_1, x_2, \ldots, x_p) \quad (10)$$

According to Sklar's 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, e.g., 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 Gaussian copulas, i.e., 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, while Frank copulas have no emphasis on tail dependency, i.e., symmetrical dependencies on both tails (Schmidt 2006). The product copula, the simplest copula, indicates independence between 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 copula parameters that were estimated using MLE. The Gaussian copula parameter was the covariance matrix. The t copula used the same matrix plus the number of degrees of freedom. The Gumbel, Clayton and Frank copulas each used a correlation parameter.

2.2.7.2 Simulated Joint Distributions

Simulations tested the GOF of the fitted copulas. These used the uniform random variables and fitted parameters for each copula (described above), as well as marginal distributions fitted for each VOC. A large number (n = 1,000) of pseudo-observations were generated for each mixture. Using the pseudo-observations, the probabilities that all components in the mixture exceeded 50^{th}, 75^{th}, 90^{th} and 95^{th} percentile cutoffs were calculated and compared to observations. For comparison, we also calculated probability assuming independence among mixture components, e.g., the probability of a three component mixture in which each component exceeded the 90^{th} percentile concentration is 0.001 ($p = 0.1^3$). Because styrene and TCE had low detection frequencies (49 and 31%, respectively), probabilities that all mixture components exceeded the 50^{th} percentile cannot be calculated.

To examine the influence of each mixture component and any trends that might be associated with concentration, mixture fractions, which were defined as a component's fractional contribution to the total concentration of the mixture, were calculated for both observed and simulated data, and results were summarized using the median fraction in several bins (50 - 75^{th}, 75 - 90^{th}, 90 - 95^{th}, 95 - 100^{th} percentile) for each mixture. Changes in the mixture fraction 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 "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 (US EPA 2000a). We also computed the fraction of individuals with cumulative risks exceeding thresholds of 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} and 10^{-2}, and compared results obtained using the observations, copula simulations, and

multivariate lognormal distributions using the observed means and variance/covariance matrix. 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).

2.2.8 Time and Exposure Fractions

The sampling time and time spent in different locations (outdoors in neighborhood, outdoors out of neighborhood, indoors at home, indoors at school/work, other indoors, transportation, and unknown) were calculated for each participant. Participants who had missing-time fractions $F_{t,miss}$, exceeding 0.25 (n = 50), were excluded. The mixing time fraction was calculated as:

$$F_{t,miss} = (T_{total} - T_{outdoor} - T_{indoor} - T_{transit})/T_{total} \quad (11)$$

where T_{total} = total time spent (min), $T_{outdoor}$ = time spent outdoors (min), T_{indoor} = time spent indoors (min), and $T_{transit}$ = time spent in transit (min).

An individual's total, cumulative or potential exposure is often represented as the sum of the concentration-time product across all compartments or microenvironments in a given time period. From the RIOPA dataset, the fraction of exposure attributable to the outdoor microenvironments was calculated for each participant as

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

where $F_{outdoor}$ = fraction of personal exposure originating outdoors in participant's neighborhood, $C_{outdoor}$ = residential outdoor VOC concentration ($\mu g\ m^{-3}$), $T_{neighborhood}$ = time spent outdoors in neighborhood (min), and $C_{personal}$ = personal VOC exposure ($\mu g\ m^{-3}$). Similarly, the indoor exposure fraction is

$$F_{home} = (C_{home}\, T_{home})/(C_{personal}\, T_{total}) \quad (13)$$

where F_{home} = fraction of personal exposure originating indoors at home for each VOC, C_{home} = indoor VOC concentration ($\mu g\ m^{-3}$) at home, and T_{home} = time (min) spent indoors at home.

These exposure fractions, which consider only two types of locations (indoors and outdoors), were computed for each VOC and participant. They assume that VOC measurements were representative of the location and error-free, and that the time-activity data were complete ($F_{t,miss} = 0$) and error-free.

With the (strong) assumptions just stated,

$$1 = F_{outdoor} + F_{indoor} + F_{other} \quad (14)$$

where F_{other} = is the exposure fraction in all other compartments, e.g., commuting and workplace. If some time is unaccounted for (e.g.,$F_{t,miss}>0$), then $F_{outdoor} + F_{home} < 1$. As discussed later, $F_{outdoor}$ was generally very small. However, $F_{indoor} >1$ for 11 to 20% of the observations (n= 52 to 98, depending on the VOC), $F_{home} >1.25$ for 5 to 11% of the data (n= 25 to 53), and $F_{home} >1.5$ for 2 to 8% of the observations (n= 11 to 39). Clearly, these cases did not satisfy the assumptions stated, i.e., the indoor time-concentration product exceeded the total personal exposure. Violation of any of the assumptions could cause such results. Considering the VOC measurement errors alone, most sampling programs set performance criteria at about 25%, and it is reasonable that roughly 10% of the measurements had greater errors. Given the importance of the indoor environment to VOC exposure, sampling error alone might explain a good fraction of the divergence from the assumptions. While cases where $F_{home} >1$ might be excluded, it seems likely that indoor exposure was important and dominant, and thus might be reasonable to assume that $F_{indoor} \approx 1$ and $F_{outdoor} \approx 0$ in such cases. In the following analysis, we excluded $F_{home} > 1.25$.

A second approach to apportion exposures to measure residual compartments might estimate the total exposure E_{total} ($\mu g\ m^{-3}$ min) as:

$$E_{total} \approx C_{outdoor}\, T_{outdoor} + C_{indoor}\, T_{indoor} + C_{other}\, T_{other} \quad (15)$$

and then use this approximate value (rather than $C_{personal}\, T_{total}$) as the denominator in eqs. (12) and (13). This remains an approximate and downward-biased estimate since C_{other} was not measured, however, if the $C_{other}\, T_{other}$ product is small, errors should be small, moreover, all fractions are sure to be less than 1. F_{home} calculated using eq. (15) was very near one, e.g., means and medians ranged from 0.96 to 1 for all VOCs, and the 75th percentile exposure fractions were 1 for all VOCs, again showing the dominance of indoor exposures. Thus, the

former method (eq. 13) was used.

Exposure fractions were stratified by city and by warm (May to October) and cool (November to April) seasons. The significance of differences was evaluated using Kruskal-Wallis (K-W) tests.

2.2.9 Identification of Determinants

2.2.9.1 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 measurements as dependent variables. These models used six VOCs (benzene, toluene, MTBE, 1,4-DCB, PERC and chloroform), which were selected to represent a range of VOCs and potential emission sources. Next, variables that attained statistical significance ($p < 0.05$) were used in forward stepwise multivariate regression models with selection based on the Schwarz Bayesian Information Criterion. While this reduced the number of variables, the resulting parameter estimates are approximate since these models do not account for possible correlations due to clustering and nesting, e.g., two seasonal samples for most participants.

2.2.9.2 Linear Mixed-Effect Models

LMMs that incorporated fixed and random effects and repeated measures (Krueger and Tian 2004) were estimated for outdoor, indoor and personal measurements 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 were evaluated. However, few significant interactions between determinants of VOC exposures were found. Thus, interaction terms were not retained in the final models. Using log-transformed VOC concentrations, random intercepts, nested effects for city, and interactions, the LMMs are expressed as:

$$\log(C_{ti}) = (\beta_0 + b_{0i}) + \beta_1 \text{Visit}_t + \beta_2 \text{City} + \ldots + \beta_n X_n + \varepsilon_{ti} \quad (16)$$

where C_{ti} = VOC concentration ($\mu g\ m^{-3}$) at time t for individual i, β = model coefficients for fixed effects, b = random deviation from the overall fixed effects, Visit_t = sample collected at

time t, X = other covariates, and ε_{ti} = random error of the VOC concentrations from the predicted line at time t for individual i. Since the LMMs used log-transformed VOCs, the effect size for each explanatory variable was calculated as follows,

$$\text{Effect size} = e^{(\beta U)} \quad (17)$$

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

To maintain a sufficient sample size, variables with fewer than 400 observations were not 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).

2.2.9.3 Model Assessment

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

$$R^2 = (\sigma_{int}^2 - \sigma_{full}^2) / \sigma_{int}^2 \quad (18)$$

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

2.2.9.4 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 were compared to the original dataset (with missing data). Three models for each sample type were selected for this comparison: models with the least missing data (e.g., 3% missing for

personal measurements of styrene), models with a modest amount of missing data (e.g., 20% missing for benzene), and models with a high amount of missing data (e.g., 28% missing for d-limonene). Differences between the original and MI datasets were computed as the relative change in model estimates of β. The results of this comparison (Supplemental Tables S3 to S5) demonstrated that while models using imputed data tended to have smaller (more statistically significant) p-values, changes 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 one (outdoor benzene) had three parameters change by more than 30%. Because missing data did not greatly affect the LMM results, subsequent results do 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).

CHAPTER 3

Results and Discussion

The results of the statistical analyses described in Chapter 2 are presented in this chapter. It includes nine sections in the order of the research objectives (Section 1.3). Section 3.1 provides descriptive statistics for the RIOPA and NHANES datasets. Section 3.2 shows the full distributions of the observed VOC data. Section 3.3 presents extreme value analyses for the VOC exposures. Section 3.4 addresses the mixture distributions for the exposures. Section 3.5 presents the trends of VOC exposure from 1988 to 2004. Section 3.6 shows the VOC mixtures identified by PMF analysis. Section 3.7 describes the dependencies and joint distributions of VOC mixtures. Section 3.8 presents time fractions and VOC exposure fractions. Section 3.9 addresses potential determinants of personal, home and outdoor VOCs. Each section (except Sections 3.1 and 3.2) also compares results with previous studies, and discusses the strengths and limitations of the analyses.

3.1 Descriptive Statistics

3.1.1 RIOPA Study

Descriptive statistics for RIOPA VOCs are shown in Table 3 to 10, and Spearman rank correlations between pollutants are shown in Table 11 to 14. Findings from these initial analyses include:

- Detection frequencies varied widely and depended on the compound. For VOCs, detection frequencies ranged from 6 to 97% for outdoor measurements; from 25 to 95% for indoor measurements; from 31 to 96% for personal adult measurements; and from 23 to 97% for personal child measurements). For $PM_{2.5}$, all of the measurements were above the MDL. One-half of the MDL was substituted for measurements below MDLs.
- For most VOCs, mean concentrations were ranked as roughly: indoor = personal > outdoor. However, for 1,4-DCB, the maximum indoor concentration (4051 μg m^{-3}) was

twice that of the highest personal concentration. The highest mean outdoor, indoor, personal adult, and personal child VOC concentrations occurred in Los Angeles, Houston, Houston and Elizabeth, respectively.

- The correlation coefficients among outdoor VOCs were generally higher and more commonly statistically significant than among indoor and personal VOCs, and there were more statistically significant pairs among the outdoor measurements.

3.1.2 NHANES III and 1999-2004

Table 15 breaks out descriptive summary statistics for the NHANES III (1988-1994) and continuous NHANES (1999-2004) cohorts. (Supplemental Tables S6 gives cohort-specific statistics.) CTC and TCE had very low DFs (5.5 and 4.8%, respectively), and were excluded from further analyses. In NHANES III, 1,4-DCB had the highest mean level (1.11 ± 0.12 µg L^{-1}) among the 12 VOCs, over twice that seen for the next highest compound, toluene, while BDCM had the lowest mean (0.008 ± 0.001 µg L^{-1}) with 86% of measurements fell below the MDL. In continuous NHANES, 1,4-DCB levels decreased (0.87 ± 0.10 µg L^{-1}), although it remained the single highest VOC. Again, DBCM had the lowest concentration (0.002 ± 0.000 µg L^{-1}) with 43% of measurements below the MDL (which also decreased). VOC levels decreased over these two periods, and differences in high-end exposures were particularly striking (Table 15). Again examining 1,4-DCB, the maximum was 52 µg L^{-1} and the 1988-1994 95^{th} percentile concentration was 11 µg L^{-1}, well above any other VOC. As discussed later, products containing 1,4-DCB have been widely used indoors, and possible occupational exposure and low clearance rates for this VOC may increase exposures and concentrations in blood.

As expected, related VOCs were correlated. The five BTEX compounds in blood had Spearman rank correlation coefficients from 0.14 (benzene and m,p-xylene) to 0.81 (ethylbenzene and o-xylene) in NHANES III, and from 0.38 (benzene and m,p-xylene) to 0.89 (ethylbenzene and o-xylene) in continuous NHANES (Table 16). The THM compounds were significantly correlated, except for chloroform and bromoform in NHANES III. In general, correlation coefficients were lower in the 1988-1994 cohorts, in part due to the higher MDLs obtained during this period.

Correlation coefficients between blood and personal air measurements in the 1999/2000

cohort were statistically significant for the nine VOCs available, and ranged from 0.24 to 0.38 for the BTEX compounds, to 0.62 for PERC and 0.65 for 1,4-DCB (Table 17). Thus, the personal air measurements explained a modest portion of the blood measurements. The NHANES study design likely lowered these correlations since sequential, rather than simultaneous, measurements were utilized, i.e., higher agreement likely would have occurred if blood was sampled when the personal air samplers were returned. Also, correlations are lowered by clearance rates that differ among VOCs, exposure pathways other than inhalation (e.g., consumption of chlorinated water), and experimental errors. Nonetheless, the positive and significant correlation suggests that the blood measurements provide useful exposure information.

Due to relatively rapid clearance, VOCs measurements in blood reflect exposures over only the immediate period preceding the blood draw (e.g., 2 or 3 half-lives). If sampling was random, blood measurements can reflect chronic exposures, although some attenuation is expected since blood draws would not immediately follow high exposure events due to time needed for travel and processing in the MEC. Consequently, the sample variability may not reflect the true variability of chronic exposures.

The 1988-1991 cohort had an excessive fraction (63%) of values reported as "extreme or illogical values" for toluene, ethylbenzene, o-xylene, styrene, bromoform and PERC, which left fewer than 200 valid measurements. Also, compared to subsequent cohorts, available data for these VOCs and cohort tended to have lower correlation among related compounds, and means (and medians) appeared inconsistent (Supplemental Tables S6). For example, m,p-xylene measurements in this cohort were very low and inconsistent with data in subsequent cohorts. Measurements of these seven VOCs in the 1988/1991 cohort were not considered to be reliable, and thus were omitted from subsequent analyses, along with the derived BTEX and ΣTHM variables. Other assessments of VOC data quality in the NHANES documentation or general literature have not been identified.

3.2 Full Distributions for VOC Observations

Table 18 shows the distribution types providing the highest GOF, based on A-D tests, by VOC measurement type (outdoor, indoor, adult personal, child personal) in RIOPA. Data were right skewed, as expected, and the most common distribution for the RIOPA VOCs was

the Pearson type 5 (right-skewed).

The nature of the VOC distributions in RIOPA can also be visualized in Figures 1 to 4 for four VOCs that often represent different sources: benzene, 1,4-DCB, PERC, and chloroform. The left-hand panels of each figure show histograms and fitted distributions; the right-hand panels show log-transformed data and distributions fitted to the transformed data. This analysis shows several features. In addition to the right skew of the data, log-transformed data show departures from normality, primarily due to two features at either end of the distribution. First, each of the VOCs show a large number of low concentration measurements, a result of setting concentrations below the MDL, which are typically addressed by setting values to one-half MDL or some similar value. As presented in Section 3.1.1, outdoor VOCs, including styrene, 1,4-DCB, MC, TCE, chloroform, d-limonene, α-pinene, and β-pinene, and indoor MC and TCE, and child measures of TCE, all had especially low detection frequencies (< 30%, i.e., most values were below MDLs). This characteristic, an artifact in the sense that it is a result of the VOC sampling and analysis method employed in RIOPA, can influence distribution fitting and data interpretation.

Figures 1 to 4 also point out show positive skew after log transformation and (remaining high) outliers that cause deviations among the upper tails of the distributions. This was especially apparent for outdoor 1,4-DCB, indoor 1,4-DCB and d-limonene, adult 1,4-DCB, chloroform, d-limonene, and PERC, and child 1,4-DCB and d-limonene. In this research, the highest values are of key interest given that these portray the highest exposures.

The full range distributions of VOCs in RIOPA and NHANES shared some similarities. Distributions were right-skewed, and the top ranked distributions for the NHANES VOCs were usually lognormal (except for MTBE, 1,4-DCB and TCE). In contrast, of the RIOPA VOCs, the top ranked distribution was lognormal for only two VOCs (PERC and chloroform). Of course, several distributions can provide quite similar fits. As examples, Figure 5 contrasts observed and modeled distributions for benzene, 1,4-DCB, PERC, and chloroform, which can be compared to the personal adult distributions shown earlier in Figures 1 to 4. This analysis showed a number of differences. First, as can be seen on the figures, the NHANES data tended not to show a mode that was attributable to measurements below MDLs. Second, measures of central tendency and other properties tended to vary. For example, NHANES

and RIOPA had median concentrations of only one VOC, PERC, that were not different (Mann-Whitney tests, $p < 0.05$); average concentrations were not different for only three compounds (1,4-DCB, PERC and chloroform, t test, $p < 0.05$).

3.3 Extreme Value Analyses for VOC Exposures

3.3.1 Predicted Health Risks for Extreme VOC Exposures in RIOPA

Estimates of individual excess lifetime cancer risks for the median, 90th and 95th percentile concentrations are shown in Table 19 (Additional statistics are shown in Supplemental Table S7). Using median concentrations, chloroform, 1,4-DCB and benzene presented the highest (and very similar) risks, 2.0 to 2.9 x 10^{-5}, respectively; risks for other VOCs were below 10^{-5}. For the 95th percentile concentrations, the same three VOCs also presented the highest risks, 1.5 x 10^{-4}, 3.6 x 10^{-3} and 7.7 x 10^{-5}, respectively; risks above 10^{-5} are also caused by ethylbenzene, MTBE, styrene, PERC and CTC. Among the RIOPA VOCs, 1,4-DCB presented the greatest risks, e.g., for the top 10% extrema, all individuals had risks exceeding 10^{-4}, 88% exceeded 10^{-3}, and 13% exceeded 10^{-2}, a high level. Additionally, 1,4-DCB's share of the total carcinogenic risk (the sum of risks across individual VOCs) increased greatly at higher percentiles, e.g., 1,4-DCB represented 17% of the total risk using median concentrations, 81% using 90th percentile concentrations, and 98% using 95th percentile concentrations. 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 also are shown in Table 19. For hematopoietic toxicity, the median and 95th percentile risks were 7.6 x 10^{-5} and 3.7 x 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 liver and renal toxicity, the median and 95th percentile risks were 1.1 x 10^{-4} and 3.7 x 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. They include only a subset of VOCs among those known or suspected to be toxicants, e.g., RIOPA did not include naphthalene, which is associated with anemia (ATSDR 2005b), or include reliable measurements of 1,3-butadiene, which is

associated with blood and lymphatic system cancers (ATSDR 2009). The two personal exposure measurements averaged together for each RIOPA participant may not be a robust measure of lifetime average exposure. 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 hr) can be higher and could possibly exceed RfC or other guidance levels for acute effects.

3.3.2 Gumbel Distributions for the RIOPA and NHANES Data

Figures 6 to 9 display model fits to the data for indoor, outdoor and personal concentrations for the same four VOCs in RIOPA discussed earlier. Table 20 summarizes results for all VOCs and sample types.

- In all cases, Gumbel distributions provided a higher fit to extrema when defined as values above the 95th percentile as compared to above the 90th percentile, suggesting that this is a more appropriate cut-off. Thus, the remainder of this analysis uses this higher cut-off.
- Higher fits ($R^2 > 0.85$) were seen for outdoor measurements of benzene, toluene, MTBE, d-limonene and α-pinene; indoor measurements of BTEX compounds, MTBE, styrene, 1,4-DCB, chloroform, α-pinene and β-pinene; personal adult measurements of ethylbenzene, m,p-xylene, o-xylene, styrene, 1,4-DCB and β-pinene; and personal child measurements of styrene, 1,4-DCB, α-pinene and β-pinene.
- Lower fits ($R^2 < 0.6$) were seen for many outdoor measurements of ethylbenzene, o-xylene, styrene, 1,4-DCB, MC, TCE, PERC, chloroform, CTC, α-pinene and β-pinene.
- Often, child personal measurements had lower fits, possibly a result of lower sample sizes which did not capture many "true" outliers.
- High fits were seen for indoor and personal measurements for several VOCs, including the BTEX compounds, styrene, 1,4-DCB, chloroform and β-pinene.
- Several VOCs did not show high fits for any sample types, e.g., MC, PERC and CTC.
- In a number of cases, an even higher cut-off might be appropriate when fitting Gumbel-type distributions, and sometimes results are driven by a few outliers.

These results suggest that simple parametric distributions do not fit the entire range of

observations in the RIOPA VOC dataset, that extreme value distributions often can provide good fits the highest values, e.g., the top 5% of measurements, and that some additional work to explore the sensitivity to cut-offs could be useful.

Although the extreme value analysis is descriptive and cannot suggest underlying causes, it does suggests that extreme values are more likely for certain VOCs and certain types of exposure measures, e.g., high personal exposures to BTEX may be associated with vehicle refueling events, high indoor levels of pinene may be associated with cleaning events, etc. For some VOCs and certain exposure compartments, outliers are unlikely, e.g., CTC is a long lived VOC with few localized sources, and other solvents and some other VOCs also have few strong and localized outdoor sources likely to produce extrema.

RIOPA and NHANES show the contrast between extreme value distributions. Most VOCs in NHANES showed better fits (higher R^2) to the maximum Gumbel distribution than the RIOPA data, although BTEX compounds showed high R^2 values in both data sets. Chlorinated hydrocarbons (TCE, PERC and chloroform) had better fits in NHANES, the opposite for 1,4-DCB. Several large differences were seen in maxima in that RIOPA had higher maximum concentrations, sometimes by very large amounts, e.g., PERC and chloroform maxima in RIOPA were 2,618 and 1,224 $\mu g\ m^{-3}$, respectively, compared to 659 and 54 $\mu g\ m^{-3}$ in NHANES. Like other compounds, maximum Gumbel distributions provided a better fit to these two VOCs in the NHANES dataset than obtained for RIOPA.

Different sampling designs and sample bias likely explain some of the differences between RIOPA and NHANES. Designed as a nationally representative sample, NHANES should reflect population heterogeneity, and if this applies to VOCs and their extrema, then NHANES should better represent the true extreme value distributions than the more stratified sampling design used in RIOPA. A second reason is protocol differences. In NHANES, staging was extensive, and included two trips by participants, in most cases by private vehicle, to a centrally-located MEC, which consisted of multiple trailers in a parking lot used for surveys, blood collection, VOC sampler deployment, and other purposes. RIOPA used in-home measurements and did not require common staging and the associated trips. This might have produced greater uniformity in the NHANES data, among other differences. We have noted discrepancies in some of the NHANES blood VOC data in earlier cohorts and only

modest correlation between VOC measurements in blood and personal air in a subset of the 1999-2000 NHANES cohort (Su et al. 2011), however, these issues are not expected to adversely affect the comparability of the air samples.

3.3.3 Generalized Extreme Value Distributions for the RIOPA Data

Table 21 shows parameters of GEV distributions fitted to the VOC data, and goodness-of-fit statistics. Figure 10 shows cumulative distributions of cancer risks for four VOCs for simulated data matching GEV, Gumbel and lognormal distributions, as well as the observed data. Separate plots are shown for the top 5 and 10% extrema. The GEV distributions closely fitted both the top 5 and 10% of observations of all VOCs based on A-D tests (Table 21), and comparisons of simulated and observed distributions matched based on K-S tests, with the exception of the top 10% of β-pinene (Table 22). With the exception of the top 5% of benzene concentrations, the shape parameters of the GEV distribution were close to or larger than 0, indicating Gumbel or Fréchet distributions, and the location and scale parameters reflected the high percentile concentrations shown earlier (Table 21). While the GEV distributions closely fitted the extrema, including 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 right-hand tails, e.g., values above the 98^{th} or 99^{th} percentile.

Gumbel distributions fitted several of the VOCs (e.g., top 5 and 10% of benzene, ethylbenzene, MTBE, styrene, 1,4-DCB, PERC and chloroform concentrations), based on K-S tests (Table 22). Sometimes the lowest values (i.e., the left tail) were lower than observations, and some values even went negative (The plots in Figure 10 are truncated and do not make this visible.)

Lognormal distributions fitted extrema for several VOCs (e.g., top 10% of benzene and ethylbenzene observations, the top 5 and 10% of MTBE, PERC and chloroform, and the top 5% of CTC, shown in Table 22. However, these distributions typically diverged from observations, and the "fat" right-hand tails were greatly unrepresented (Figure 10). We note

that the lognormal distributions were fitted for the full dataset, not just the top 5 and 10% used for the GEV and Gumbel distributions.

The observed and predicted fraction of individuals with risks that exceed 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} and 10^{-2}, risk cut-offs that might be considered "bright lines", are examined in Table 23. This analysis is performed for the top 5% and 10% of the data, and the three distributions. GEV and Gumbel predictions were very close to observed frequencies, and differences were usually within a few percent. As an example, for the top 10% of the benzene data, the observed, GEV, Gumbel and lognormal simulations showed 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, the corresponding frequencies were 25%, 27%, 24% and 10%. As noted earlier, GEV simulations sometimes overpredicted the very highest upper percentiles (seen at the 10^{-4} risk level for ethylbenzene, MTBE, styrene, TCE, PERC, chloroform and CTC), and such risks were not seen in the data. However, such cases were rare, comprising less than about 1% of the entire dataset. Gumbel distributions also overpredicted extrema (although maxima were lower), and also underpredicted lower risks, in part due to its unbounded nature that can generate small and negative values. For example, all (100%) observed individuals had risks exceeding 10^{-6} for MTBE, styrene, 1,4-DCB, TCE, PERC and CTC, but Gumbel predictions ranged from 77% (TCE) to 99% (MTBE). As noted above, lognormal predictions did not match observations, and the differences could be large, e.g., 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 3-parameter GEV distributions provided better fit than the 2-parameter Gumbel distribution. In contrast, lognormal distributions provided poor fits to extrema.

3.3.4 Generalized Extreme Value Distributions for the NHANES Data

In most cases, the top 5% and top 10% of the NHANES data did not match GEV distributions fitted to either the larger dataset, which used sample weights to specify repeat

frequencies, or to the smaller (equal size) datasets that used bootstrap methods and repeated sampling (Table 24 and Table 25). 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 test. Possibly the two approaches used to incorporate the sampling weights did not decrease the "staircase" nature of the weighted datasets, which 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 should be drawn from a set of independent, identically distributed samples (Fisher and Tippett 1928). We tried a third approach, fitting GEV distributions to the unweighted NHANES data, which did match on basis of A-D and K-S tests (Table 26). These results suggest that the fitting or possibly the evaluation approaches used for the GEV distributions are inappropriate for weighted datasets.

3.3.5 Limitations

This work has several limitations. GEV and Gumbel distributions describe only one tail of a distribution, and cannot be used for the remainder of the distribution. Cancer risk estimates require long-term exposure estimates, and averaging the two visits in the RIOPA dataset may not be representative of long-term exposure. Additionally, individuals lacking either data from either visit were excluded, which reduced the sample size. Extrema were defined using two cut-offs (90 and 95th percentiles). The use of a higher cut-off, e.g., the 98th percentile, was not feasible due to sample size issues. The results for RIOPA are limited to personal exposure measurements of 15 VOCs made in three large cities in the USA. Because RIOPA included only non-smoking households, and for other reasons noted earlier, its results are not generalizable to other cities. We did not evaluate extreme value distributions for other VOCs (e.g., formaldehyde) or other pollutants (e.g., $PM_{2.5}$). There may be additional explanations for the differences between the RIOPA and NHANES results beyond those noted (i.e., different sampling designs, staging, demographics, and presence of smokers).

3.4 Mixture of Normal Distributions for VOC Observations in RIOPA

3.4.1 Single Normal Distributions

For chloroform, which is roughly lognormally distributed except that 17% of the data is under the MDL, the single normal distribution model fits about as well as the finite mixture of

normals and DPM of normals (described below) on the basis of MSE and MAE values, and gives a 21% probability of being below the MDL, similar to that observed (Table 27). However, for 1,4-DCB and styrene, which have more data under the MDL as well as heavy tails, the fit of the single normal distribution model is inferior compared to those of the mixture models. For example, the predicted probability of being below MDL is 28% and 56% for 1,4-DCB and styrene, respectively, compared to 34% and 66% observed, and 33% and 64% estimated by the mixture models. The single normal distribution overestimated the mean of these VOCs since it underestimated the non-detection frequency.

3.4.2 Finite Mixture of Normals

Fitted density plots (and component clusters) are shown in Figures 11B, 12B and 13B for chloroform, 1,4-DCB and styrene, respectively. The fitted parameters (weight λ_k, location μ_k and dispersion $\sigma_k{}^2$) of each cluster K for the mixture of normals are given in Table 28. The optimal Ks (based on the AICc) were 2, 4 and 3 for chloroform, 1,4-DCB and styrene, respectively. These choices of K clearly reflected the multi-modality and right skewness of the VOC data, and the resulting mixture of normals closely fitted the observed distributions. For example, Figure 12B represents the four clusters that fitted the 1,4-DCB data: the first (red) cluster captured the left censoring due to the MDL, the second and third (green and blue) clusters reflected the majority of the data and the skewness, and the fourth (blue) cluster modeled the heavy tail.

3.4.3 Nonparametric DPM of Normals

Fitted densities using DPM of normals for the three VOCs are shown in Figures 11C, 12C and 13C. This method clearly captures the censoring, right-skewness, and potential multi-modality of the exposure data. In terms of MSE and MAE, the DPM approach attained slightly lower values than the finite mixture of normals (Table 27).

Panel D on Figures 11 to 13 show results of the sensitivity analysis with the four different gamma distributions used as priors for precision parameter α. As noted before, K stochastically increases with α as $\mathrm{E}(\mathrm{K}\,|\alpha, \mathrm{n}) \approx \alpha \log(1 + \mathrm{n}/\alpha)$ for moderately large n (Antoniak 1974). The four prior distributions were informative and formed up to 20 clusters that reflected more specific subject matter information. Estimated densities obtained using the four priors nearly overlapped and showed very similar MSE and MAE for each of the

VOCs, although the corresponding posterior distribution of the number of clusters K varied (Table 29). The posterior mean of K under all prior settings of α (Table 29) slightly exceeded the K selected using the AICc (Table 28). The higher K in the DPM is due to the prior information of α, and does not introduce any additional complexity or more model parameters. The initial prior variance of α critically influences the extent of smoothing (Escobar and West 1995). 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 decreased 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 (8) is data driven. As shown in (8), the DPM can handle values under 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 assume that observed data randomly arise from sub-distributions with certain probabilities as the finite mixture of distribution models. (Again, sub-populations that an individual observation belongs are not identified.) Compared to 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 on priors for precision parameter α. The density estimation results were robust given a moderate sample size (n = 544) without any convergence issues noted.

3.4.4 Simulations

Simulation results, summarized in Table 30, show similar patterns for the MSE and MAE criteria. Both finite mixture and DPM of normals provided much better fits than a single normal distribution, except that the former two methods are only slightly better under distribution 1 with $P_0 = 0.15$. For both distributions, as the fraction P_0 of data below the

MDL increased, there is evidence of increasing trend of lack of fit for a single normal distribution, while the finite mixture and DPM of normals fitted considerable better and without such trend. The DPM of normals shows advantage of robustness regarding P_0. It fits equally well, or even better, as P_0 increased. For distribution 1, the finite mixture of normals provided a slightly better fit than the DPM of normals, but this trend can be offset since the prior variance of α can be decreased to promote smoothness. In this regard, DPM is much more flexible than the finite mixture of normal. Here, we have used $\alpha \sim \text{Gamma}(1.2, 2.5)$ which favors 1-5 clusters given our sample size (as the prior information of K). For distribution 2 which is right skewed and with a heavy tail, the DPM of normals provided a much better fit than finite mixture of normals 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, fat tails, 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/or 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.

3.5 VOC Trends from 1988 to 2004

Potential covariates were identified before evaluating VOC trends in NHANES cohorts. Several occupational groups were associated with VOC levels, although none achieved statistical significance in ANOVA tests, possibly because effects were small or diluted due to the broad occupational categories used. Nevertheless, trend analyses were adjusted for groups that seemed likely to have VOC exposure: service occupations (associated with elevated 1,4-DCB levels); precision production, craft and repair occupations (BTEX); and operators, fabricators, and laborers (BTEX). A variable combining these groups was used as a covariate in QR models. Additionally, all VOCs except PERC were associated with serum cotinine levels, which dropped from an average of 107 to 70 ng mL^{-1} over the 1988-2004 period. Initially, all QR models were adjusted using log-transformed cotinine levels. However, this variable was not statistically significant for non-aromatic VOCs and parameter estimates changed little, thus cotinine was maintained in the final QR models for only aromatic

VOCs. Among demographic variables, only age and education differed significantly between NHANES cohorts, and both age and college attainment increased with time. QR models including these variables showed insignificant changes in parameter estimates, and thus the demographic variables were not included in the final models.

The trend analysis focused on concentration quantiles exceeding 0.5 (50^{th} percentile). Often, lower quantiles were at or near MDL concentrations. Linear QR models representing the entire study period (1988 to 2004) and adjusted for solvent-related occupations and cotinine levels (aromatic VOCs) showed statistically significant trends at 0.5, 0.75 and 0.95 quantiles for all VOCs except for PERC at the 0.5 quantile, and styrene and 1,4-DCB at the 0.95 quantile (Table 31). For most VOCs, these changes corresponded to an average decrease of 2.5 to 6.4% per year (Table 32). Graphical interpretations of results for benzene, 1,4-DCB and PERC are presented in Figures 14 to 16. Panel A of each figure shows box plots for the five cohorts, superimposed with the estimated linear QR trend lines; panel B shows quantile plots of the linear QR estimate at 0.25, 0.5, 0.75 and 0.95 quantiles, along with 95% confidence intervals. Due to low DFs, the 0.25 quantile (left-most point) is not meaningful for 1,4-DCB and PERC, and only somewhat meaningful for benzene. These plots suggest that the rate of decline can depend on the quantile, and three patterns were discerned across the VOCs. Pattern 1 has similar decreases at all quantiles, shown by benzene (Figure 14B). This pattern suggests uniform emission and/or exposure reductions from the sources that dominate population exposures, e.g., reduced exhaust and evaporative emissions from vehicles, the largest benzene exposure source. Pattern 2 shows more rapid decreases at upper quantiles and slower decreases at lower quantiles, as seen for PERC (Figure 16B). In this case, the most exposed cohort might have a unique exposure source, which has been controlled, or that other measures have been taken to limit high exposures, while lower level exposures continue largely unabated among the general population, possibly due to other sources that have not been controlled as much. This pattern could be explained by controls on the leading occupational exposure sources of PERC, e.g., dry cleaning and metal-degreasing operations. Pattern 3 is a rapid decrease at central quantiles that exceeds upper quantiles decreases, as seen for 1,4-DCB (Figure 15B). This may result from controls on sources that affect indoor and/or outdoor concentrations, without a commensurate reduction in high exposure cases. For 1,4-DCB, this might be explained by reduced use of mothballs and air fresheners, the major

exposure sources for the general population, while the most exposed individuals either continue to experience a separate exposure source, e.g., industrial production of repellents, insecticides, resins, etc., or they remain intensive users of this chemical. Patterns and possible sources for individual VOCs are discussed in the next section.

The trend analysis also raised questions regarding the veracity of the 1999/2000 VOC data, which had the highest levels of benzene (average of 0.184±0.015 μg L^{-1}) and chloroform (0.058±0.005 μg L^{-1}) across five NHANES cohorts. Moreover, using the 1999/2000 data as a baseline, subsequent cohorts showed very rapid declines (>15% per year to 2003/2004) in median and higher percentile concentrations of benzene, toluene and chloroform, far faster than earlier years (Table 32). As noted, previous discussions of the comparability of this or other cohorts in the VOC dataset have not been seen. To investigate the sensitivity of results to the 1999/2000 cohort, linear QR models were rerun without these data. While this lessened the rate of decrease, differences were generally small, e.g., slopes changed by less than 30% for all VOCs and quartiles except benzene and toluene (0.75 quantile), BTEX (0.5 quantile), styrene (0.75 and 0.95 quantiles), chloroform and ∑THM (0.95 quantile), and few coefficients differed statistically (based on Wald tests assuming nil covariance between the two slopes) except benzene, toluene, o-xylene, and BTEX (0.5 quantile), benzene, toluene, bromoform, and PERC (0.75 quantile) (Supplemental Table S8). Bromoform and PERC at the 0.5 quantile also showed differences, but these were attributable to low DFs and are not meaningful. In summary, long-term trends were not strongly dependent on the 1999/2000 data, and thus these data were kept in subsequent analyses.

A second sensitivity analysis was undertaken that used piecewise linear QR models allowing changes in trend over the study period. As before, models were adjusted for solvent-related occupations and cotinine. QR model results using a knot at 1999/2000 are shown in Supplemental Table S9. (Knots at other locations provided poor fits.) This analysis indicates that for most VOCs, declines from 1988 through 2000 were either not statistically significant or considerably smaller than declines from 1999/2000 through 2004, and that several VOC increased over the 1988-2000 period (including benzene and chloroform at the 0.5 quantile, benzene, toluene, styrene and chloroform at the 0.75 quantile, and benzene, m,p-xylene, styrene, and chloroform at the 0.95 quantile). Declines in the second period (shown as Slope2 in Supplemental Table S9) were reasonably consistent for the aromatic

VOCs and chloroform, and faster than those from the linear QR models that spanned the entire period (Table 31). Overall, the piecewise QR models are similar to results in Table 32, and likewise suggest that reductions in blood VOC levels were largely accomplished from 1999/2000 onward. However, the piecewise models are less robust than the linear QR model since slopes for each time period use only three cohorts (or time points), and sometimes only two in the first period (1988-2000) since portions of the 1988-1991 data were omitted, and since they depend strongly on the 1999/2000 cohort data, which have several anomalies as noted previously. Moreover, trends in ambient concentrations for most VOCs do not support this steeper decline, as discussed below.

The third sensitivity analysis compared both linear and piecewise regression models with and without adjustments for strata and clusters. This showed only small differences in most cases: standard errors were larger for most VOCs, however, differences were significant for only BTEX among the linear models, and for DBCM, bromoform and PERC among the piecewise models. Although we cannot account for NHANES' cluster sampling protocol in the QR models, these results suggest that the QR model results are reliable.

In summary, VOC levels in the NHANES blood samples substantially declined over the 15 year period. While piecewise models suggest that exposures to some VOCs did not decrease in the 1990's and then rapidly declined in the early 2000's, this may be driven by anomalies in the NHANES data, as discussed below.

3.5.1 Interpretation and Reliability of Trends

Many factors can affect the interpretation and representativeness of the NHANES data. First, while each cohort was designed to be nationally representative, biases might result from unknowingly over-sampling populations that are more exposed, genetically special (e.g., unable to rapidly clear VOCs), or otherwise not representative. As noted earlier, only minor group differences were seen among the demographic variables, literature discussing biases has not been identified, and while genetic differences can affect results, the biomarker documentation does not specify any such factor that affects the interpretation of VOC measurements in blood (ACGIH 2001). Second, statistical variation is inherent in any sampling program and some cohorts had smaller PSU and sample sizes, but considering the NHANES sample sizes, this should not cause systematic biases. Third, whether the

NHANES blood measurements represent valid exposure measures could be questioned, and indeed the approximate nature of these biomarkers was indicated by only modest correlation with air samples and the rapid clearance in the blood (discussed earlier). In this case, however, a bias towards the null (no trend) would be the likely outcome, which was not seen. Fourth, changes in protocols, including the air sampling conducted in the 1999/2000 cohort, the shift from NHANES III to continuous NHANES, or some other unknown study element, could affect results. We did identify NHANES data that appears suspect, and either excluded it or used sensitivity analyses to obtain confirm interpretations. Nothing emerged that could explain observed patterns.

Several independent findings support the long-term VOC exposure trends derived from NHANES. First, the NATA emission inventory, while including only a few of the VOCs in measured in NHANES, reports that emissions of several VOCs increased in the 1990's, e.g., benzene increased from 337,000 to 410,000 tons/year from 1996 to 2002, and chloroform increased very markedly from 3,310 to 15,139 tons/year from 1996 to 1999; Table 33). Annual average ambient concentrations predicted by NATA, spatially averaged, show negligible movement from 1996 to 1999 for benzene, chloroform, PERC and 1,4-DCB, and decreases of 3.9 to 18% per year for benzene, toluene, xylene and PERC from 1999 to 2002. These data support some of the piecewise trends, and also the high levels of benzene and chloroform seen in NHANES in 1999/2000, however, exposure analyses using emission inventories have limitations, as discussed in the Introduction.

Ambient air monitoring provides a more direct exposure measure. PAMS data are summarized in Table 34. For the 2001-4 period, annual mean concentrations of benzene, toluene, ethylbenzene and o-xylene in the UATMP network decreased by 11 to 20% per year, and by 7 to 11% per year in PAMS. Thus, recent UATMP and PAMS trends are roughly similar, though UATMP concentrations are lower. Considering the older (1993-1999) PAMS data, annual mean concentrations of aromatic VOCs decreased from 4.4% per year (toluene) to 11% per year (styrene), and for five of the six VOCs measured, the rate was half that seen in the 1999-2004 period. Issues regarding the spatial and temporal coverage of PAMS data were discussed in the Introduction. The AQS data may be more revealing, and annual means of the nine VOCs common to NHANES are tabulated and plotted in Table 35. Regression analyses show approximately linear decreases of 5 to 7% per year for benzene, toluene, ethylbenzene

and styrene from about 1990 to 2004. Trend plots show comparable long-term decreases and hints of somewhat accelerated trends since 2000 for m,p-xylene, o-xylene and 1,4-DCB. Chloroform shows a dramatic 21% per year decrease from 1990 to 1994, which then shows a flat trend. PERC levels decrease by 6.7% per year, although the trend is erratic. While ambient measurements too have limitations as exposure indicators, the national-level data show that ambient concentrations of many VOCs have declined in a linearly over 15 years, and the rate appears slightly faster than those based on the NHANES exposure data. For several VOCs, some evidence suggests swifter declines after 2000, however, the ambient data does not reflect the high levels of benzene and chloroform in the 1999/2000 NHANES blood data.

In summary, ambient and emission data for most VOCs show strong downward trends from about 1990 through 2004. Regarding indoor exposures, national-level corroborating evidence is unavailable, however, there is linkage with ambient data in that outdoor concentrations represent a "floor" for indoor levels, and because the emission controls on fuels and vehicles that lower ambient VOC concentrations will also reduce exposures while commuting and in buildings with attached garages (Batterman et al., 2006). We next examine trends of individual VOCs.

3.5.2 Benzene

Over the 15 year study period, benzene exposures in NHANES declined by 3.3 to 4.3% per year, depending on the quantile. As noted, benzene trends matched pattern 1, with relatively consistent decreases at all quantiles, which parallel some of the emission and airborne concentration trends. Benzene was listed as a hazardous air pollutant by U.S. EPA in 1977 and as a carcinogen in 1986, and many emissions have been inventoried and regulated. U.S. emissions fell from 493,000 to 386,000 T yr^{-1} tons between 1990-1993 and 2005 (US EPA 2009b), representing a 1.5% per year decrease. On-road vehicle emissions, the single largest source category, declined faster, from 312,000 to 143,000 T yr^{-1} or 3.6% per year. Further restrictions of benzene content in gasoline were issued in 2007, and additional reductions in mobile source air toxics emissions (including benzene) are anticipated (US EPA 2010c). Benzene is metabolized fairly rapidly with a half-life in blood of about 8 hr (Brugnone et al. 1992).

Inhalation exposure to benzene has been extensively reviewed (ATSDR 2007a).

Ambient measurements declined by 4.5 to 4.9% per year from 1994 to 2008; medians dropped from 2.10 to 0.79 µg m^{-3}; and 90th percentile levels fell from 5.03 to 1.59 µg m^{-3} (US EPA 2009a). Urban concentrations fell faster, e.g., PAMS data show 8.4%, 7.2%, and 6.9% per year declines at 0.5, 0.75, and 0.95 quantiles from 1993 to 2004 and AQS data (Table 34). Since few indoor sources exist other than smoking, benzene concentrations in outdoor, indoor and personal air can be similar (PL Kinney et al. 2002), however, an attached garage can elevate residential levels (Batterman et al. 2006). Differences in biomarker and ambient trends are reflected by the relatively low correlation between blood and personal airborne levels (r=0.24, Table 17). Occupational exposures in many settings have substantially declined, e.g., median personal concentrations of laboratory technicians at a refinery dropped from 319 to <32 µg m^{-3} from 1977 to 2005 (Panko et al. 2009), however, national statistics on occupational exposures are unavailable. As mentioned, tobacco smoke is an important exposure source (L Wallace et al. 1987), and about 50% of benzene exposure in the U.S. has been apportioned to active and passive smoking (ATSDR 2007a). However, NHANES data continued to show declines in each quantile after cotinine adjustment. Overall, the trends suggest that reductions in population exposure, as reflected in NHANES, have been driven largely by reductions in gasoline- and vehicle-related emissions.

3.5.3 Toluene

Over the 1988 to 2004 period, toluene exposures decreased by 4.7 to 5.7% per year, depending on the quantile. Like benzene, toluene reductions fit pattern 1 (consistent decreases across quantiles), which indicates improved control of general exposures, e.g., vehicle exhaust, as well as high-concentration exposures, e.g., architectural paints, which are now limited in VOC contents to 250 and 500 g L^{-1} for flat coatings and graphic arts paints, respectively (US EPA 1998). Toluene is one of the more prevalent components associated with vehicles and, unlike benzene, many household products contain and emit toluene. NATA emissions decreased from 996,443 to 884,066 T yr^{-1} between 1999 and 2002, or 3.8% per year, on-road emissions decreased from 460,240 to 428,672 T yr^{-1}, or 2.3% per year (Table 33) (US EPA 1999a, 2002), and average ambient predictions declined from 3.0 to 2.5 µg m^{-3}, or 5.2% per year (Table 36). Ambient concentration at PAMS sites decreased by 6.4-8.5% per year, depending on quantile (Table 34), while annual means in the AQS data declined by 5.7% per year (Table 35). Like benzene, blood and airborne levels had only modest correlation (r=0.26,

Table 17). Toluene's half-life in blood is short, about 4.5 hr (Brugnone et al. 1986), thus blood levels tend to reflect current exposures.

3.5.4 Other BTEX Compounds

QR results for the remaining BTEX compounds for the 1988-2004 period showed significantly downward trends that tended to fit pattern 2 (rapid decreases at upper quantiles), even after adjustment for cotinine (Tables 31 and 32). Ethylbenzene, m,p-xylene, o-xylene, and styrene concentrations in blood decreased by 2.5 to 5.6% per year at each quantile. The composite BTEX exposure showed consistent decreases across quantiles in the same period; benzene and toluene contribute disproportionately to this indicator. The half-life of ethylbenzene in blood is very short (<1 hr) (Adams et al. 2005; ATSDR 2007b); xylenes are reported to have biphasic half-lives: 0.5-1 hr initially, followed by 20-30 hr (US EPA 2003b); and styrene has biphasic half-lives of 0.58 and 13 hr in blood (ATSDR 2007c). Thus, blood tends represent only recent exposures. Correlation coefficients between personal air and blood for ethylbenzene, m,p-xylene and o-xylene in the 1999/2000 NHANES cohort were 0.35, 0.38, and 0.36, respectively, higher than seen for benzene and toluene (Table 17).

In the NATA database, nationwide emissions of o- and m,p-xylene fell from 712,084 to 595,241 T yr^{-1} between 1999 and 2002 (Table 33), or 5.5% per year, and on-road vehicle emissions decreased from 269,500 to 247,765 T yr^{-1}, only 2.7% per year (US EPA 1999a, 2002). Ambient measurements fell faster, e.g., median levels of aromatic VOCs in PAMS fell by about 9% per year from 1993 to 2004 (Table 34), and AQS means fell by 5.8 to 6.4% per year, with faster declines after 2000 (Table 35). Thus, ambient levels fell more rapidly that the roughly 4% per year seen for NHANES blood VOC levels from 1988-2004 (Table 32), but less rapidly than the more recent (1999-2004) blood VOC data. The divergence suggests that reductions of indoor VOC sources trailed outdoor reductions by perhaps a decade.

3.5.5 THMs

Chloroform was the most prevalent THM. With the 1999/2000 data included, levels declined rapidly at upper quantiles (pattern 2), while comparable reductions of about 4% per year were seen across quantiles when comparing starting and ending cohorts (Table 31 and 32). BDCM, DBCM and bromoform showed rapidly decreases at central quantiles over the study period. Due to low DFs, trends at lower percentiles could not be evaluated (Supplemental

Table S2). Over the 15 year study period, concentrations decreased by 5.0 to 7.9% per year for the median, and by 3.0 to 7.5% per year for upper quantiles.

Exposures of individual THMs, including chloroform, are likely to be highly correlated, although this was not consistently shown in the NHANES blood measurements (Table 16). This can be explained, in part, by the rapid clearance of THMs from blood, e.g., half-lives of about 0.5 hr (Ashley and Prah 1997), and a biphasic clearance pattern is reported for chloroform with half-lives of 9 to 21 min and then 86 to 96 hr (ATSDR 1997a). Given these rates, the blood data represent only recent exposures. Chloroform showed a moderate but significant correlation (r=0.38) between blood and personal air concentrations (Table 17).

NATA emissions of chloroform jumped from 3,310 to 15,139 T yr^{-1} from 1996 to 1999, or 119% per year, followed by a decline in 2002 to 6,805 T yr^{-1}, or 18% per year (Table 33) (US EPA 1996, 1999a, 2002). The dramatic increase from 1996 to 1999 is likely due to changes in inventory procedures (US EPA 1999a). Predicted ambient concentrations increased by 0.9% per year from 1996 to 1999, and then decreased by 1.7% per year (Table 36). Interestingly but perhaps serendipitously, the period of highest chloroform emissions (1999) corresponded to the highest blood measurements in NHANES (Supplemental Table S6). In the mid-1990s, Maximum Achievable Control Technology standards limited emissions of halogenated solvents at industrial and waste treatment facilities (US EPA 2000c). About the same time, maximum contaminant levels on THMs in drinking water were imposed, which is probably the largest exposure source (both ingestion and inhalation) of THMs for the general population. (NATA estimates do not account for THM emissions in to drinking water, but the NHANES blood data does account for the ingestion pathway.) Lowering THMs in drinking water is expected to decrease levels at all quantiles (pattern 1). Ambient concentrations of chloroform show a trend unique among the VOCs: early decreases of nearly 21% per year for the 1990-1994 period, followed by a flat trend from 1995 onward (Table 35). Exposures of the brominated THMs had inconsistent trends, which is attributed to analytical uncertainties resulting from low concentrations (generally 10 times lower than chloroform).

3.5.6 Other VOCs

Styrene exposures significantly decreased at 0.5 and 0.75 quantiles, e.g., median levels fell by 3.8% per year over the study period (Table 32), but much faster (18% per year) from

1999 to 2004. Serum cotinine and blood styrene levels in NHANES were correlated (r=0.49), but QR models adjusted for cotinine levels continued to showed a declining trend (Table 31). (NATA only included styrene data in 2002.) Ambient concentrations of styrene in PAMS declined by about 8% per year, depending on quantile, over the 1993-2004 period, while AQS means declined by 5.5% per year, though the data showed considerable scatter (Appendices H and I). Styrene is used in reinforced plastics manufacturing, and indoor emissions can occur from building materials and tobacco smoke (ATSDR 2007c). It has biphasic half-lives of 0.58 and 13 hr in blood (ATSDR 2007c).

1,4-DCB decreased by 3.5% per year over the 15 year study period (Table 32). Decreases were more rapid at median quantiles, (pattern 3), and the 0.95 quantile result was not significant (Table 31). 1,4-DCB is widely used in mothballs, other pest repellents and toilet-deodorizer blocks, and airborne levels in occupational settings occasionally reach very high levels, e.g., 4,350 mg m^{-3} in a mono- and dichlorobenzene manufacturing plant (IARC 1982). In the US, mean and median indoor 1,4-DCB concentrations were 24 µg and 1.7 µg m^{-3}, respectively (ATSDR 2006b); the large difference reflects the highly skewed distribution of this VOC. A Japanese study found high indoor levels (mean = 114 µg m^{-3}), far above outdoor levels (3.4 µg m^{-3}) (Azuma et al. 2007). 1,4-DCB's half-life is estimated to be 7.1-8.1 hours in rats (no human data are available (Hissink et al. 1997; Boutonnet et al. 2004). NATA emission estimates of 1,4-DCB fell from 12,794 to 7,244 T yr^{-1} between 1999 and 2002, or 15% per year (Table 33) (US EPA 1999a, 2002). Ambient concentrations are low, and median concentrations among 11 sites declined by 5.0% per year from 1995 to 2005, and by 10% per year among 32 sites from 2000 to 2005 (McCarthy et al. 2007). Among the AQS VOCs, 1,4-DCB showed the strongest decrease after 2000 (Table 35). As noted, 1,4-DCB had the highest air-to-blood correlation coefficient among the NHANES VOCs (r=0.65, Table 17), thus exposures tend to reflect personal air concentrations.

PERC exposures declined by 3.2 to 6.4% per year, depending on quantile, over the 15 year study period, and decreases at upper quantiles were faster (pattern 2) (Table 31 to 32). PERC's half-life in blood, 12 to 16 hr (ATSDR 1997b), is the longest among the VOCs, and its air-blood correlation was relatively high (r=0.62, Table 17). NATA emissions increased by 2.0% per year, from 44,100 to 46,793 T yr^{-1}, between 1996 and 1999, followed by a 8% per year decrease to 35,613 T yr^{-1} in 2002 (Table 33) (US EPA 1996, 1999a, 2002). However,

predicted ambient concentrations decreased slightly, 1.5% per year, between 1996 and 1999, and then by 18% per year between 1999 and 2002 (Table 36). Nationwide emission data before 1993 are not available. MACT standards for dry cleaners, perhaps the major urban source of PERC (US EPA 2010c), were initiated in 1993. Although the AQS means show considerable variation (Table 35), the long term decline of ambient concentrations nearly exactly corresponds to the rate seen in blood.

3.6 Selected VOC Mixtures in RIOPA based on PMF

VOC sources are identified on the basis of the VOC composition using PMF analyses. In cases, several source types can contribute to a factor, or sources may have collinear emission profiles (source compositions) and thus cannot necessarily be distinguished. The following show the possible VOC composition on the basis of emission sources by sampling types.

3.6.1 Outdoor VOCs

Outdoors, apportionments were dominated by gasoline-related sources, and seasonal variation was observed. Results of sources apportionment of VOCs in RIOPA study are presented in Table 37. In warm season, four categories were shown: the dominant component in mixture 1 was MTBE, indicating gasoline vapor; mixture 2 mainly included BTEX & β-pinene, representing vehicle exhaust and biogenic sources; mixture 3 was dominated by d-limonene, representing some odorants; mixture 4 contained TCE, PERC and α-pinene which may be from industrial emissions and biogenic sources. In cold season, there were four groups: mixture 1 mainly contained BTEX compounds, indicating vehicle exhaust; mixture 2, like mixture 1 in warm season, was dominated by MTBE, representing gasoline vapor; a lot of VOCs were included in mixture 3, e.g., 1,4-DCB, TCE, CTC, d-limonene, α-pinene and β-pinene, which may come from industrial emissions; PERC, the dominant VOC in mixture 4, was used in dry cleaning industry. Gasoline-related sources (more than 60% of the contributions) were prevailing for outdoor VOCs in both seasons.

Figure 17 presents the median ratios of four common VOC groups, including aromatics, MTBE, chlorocarbons, and terpenes, by quintiles of TVOC concentrations to show VOC composition at different levels. For all outdoor VOC observations, aromatics, including benzene, toluene, ethylbenzene, m,p-xylene, o-xylene and styrene, were less abundant in the

2nd and 3rd quintiles, and MTBE was more abundant in middle and highest quintiles. The gasoline-related VOCs showed more abundance by quintiles. In contrast, chlorocarbons, including 1,4-DCB, MC, TCE, PERC, CTC, and chloroform, and terpenes, including d-limonene, α-pinene, and β-pinene, showed less abundance in higher quintiles. In the first quintile, 15% of TVOC was terpenes, and then the abundance dropped to 5% in the last quintile. Outdoor terpenes were emitted from biogenic sources, representing relatively stable background levels. Higher concentrations of TVOC may mainly attribute to other VOCs from anthropogenic sources. Thus, terpenes' abundance decreased in high quintile due to increases of other VOC concentrations. VOC measurements in different cities and seasons showed similar abundance with overall measurements, except for samples in Houston, which have more abundance of MTBE in higher quintiles.

3.6.2 Indoor VOCs

Indoor apportionments in warm and cold seasons were similar, and cleaning products and odorants were the major sources. There were four common factors for indoor VOCs in both seasons (Table 37): mixture 1 was dominated by 1,4-DCB, indicating moth repellents and odorants; mixture 2 contained d-limonene, α-pinene and β-pinene, representing cleaning products and air fresheners; mixture 3 mainly contained aromatics, TCE, PERC, chloroform and CTC, which may come from vehicle exhaust and chlorinated solvents using for degreasing; MTBE was the dominant compound in mixture 4, and indicated gasoline vapor. Cleaning products and odorants were the leading emission sources for indoor VOCs in both warm (73% of the contributions) and cold (66% of the contribution) seasons.

Aromatics and MTBE showed less abundance in higher quintiles for indoor VOCs (Figure 18). Abundance of gasoline-related VOCs in the 5th quintile was about 16% comparing to 44% in the 1st quintile, and there was no difference between warm and cold seasons. Indoor gasoline-related VOCs are mainly generated by outdoor sources, and affected by transportation and penetration process. Other VOCs, e.g., 1,4-DCB and d-limonene, generated by indoor sources, have extreme values to lead to large proportion of abundance in the higher quintiles. For example, the average concentration of 1,4-DCB in 4th quintile of TVOC was 10 μg m^{-3}, indicating 1.8% of median abundance, and the average in 5th quintile was 327 μg m^{-3}, indicating 27% of median abundance. Similar pattern was

observed for d-limonene in higher quintiles. Variations of VOC abundance were shown among cities, especially in Houston. In Houston, a quarter of 1,4-DCB samples in 5th quintile were above 1000 μg m^{-3} (only one 1,4-DCB sample was above 1000 μg m^{-3} in Los Angeles and Elizabeth).

3.6.3 Personal VOCs Consisting of Adult and Child Measurements

Dominant VOC sources for personal exposures were cleaning products and odorants, and seasonal effects were also observed (Table 37). In warm season, four groups of VOCs were shown: mixture 1, including d-limonene, α-pinene and β-pinene, indicated the use of cleaning products and odorants; ethylbenzene, m,p-xylene and o-xylene in mixture 2 represented motor sources; benzene and MTBE contained in mixture 3 indicated gasoline vapor; mixture 4 containing 1,4-DCB, TCE, PERC, chloroform and CTC suggested exposures to moth repellents and chlorinated solvents. In cold season, VOC apportionments were still dominated by cleaning products and odorants, like d-limonene, α-pinene and β-pinene (more than 40% of the contributions in both seasons). The other three VOC groups included: mixture 2 (benzene, toluene, MTBE, styrene, 1,4-DCB, TCE, chloroform and CTC) indicating gasoline, chlorinated solvents, and cleaning products, mixture 3 (ethylbenzene, m,p-xylene and o-xylene) representing vehicle exhaust, and mixture 4 (PERC) from dry cleaning solvent.

Like indoor VOCs, gasoline-related VOCs were less abundant in higher quintiles with variations between cities (Figure 19). Personal samples showed more abundance of chlorocarbons in the highest quintile than indoor samples, suggesting that people contacted the emission source, e.g., moth repellents, directly or extensively. For example, the median concentrations of 1,4-DCB in the highest quintiles were 65 μg m^{-3} for indoor samples, and 95 μg m^{-3} for personal samples. No significant differences of abundance between seasons were found. However, large variations were observed among cities, especially in Houston. Chlorocarbons were the majority (85%) in highest quintile in Houston, and other VOC groups were less than 10%. On the other hand, aromatics and terpenes were dominant in the highest quintiles in Los Angeles and Elizabeth. It was because most extreme values of 1,4-DCB were measured in Houston. Eighteen out of 66 1,4-DCB measurements were above 1000 μg m^{-3} in Houston, but there were only two measurements in Elizabeth over that value, and none in Los Angeles. Thus, extreme values of chlorocarbons in the highest

quintiles resulted in less abundance of other VOC groups in Houston.

3.6.4 Personal Adult VOCs at the First Visits

Based on the PMF analysis, four VOC mixtures were identified (and designated as mixtures A1 to A4 in Table 38 and Figure 20):

- Mixture A1 contained benzene (average contribution = 1.4 μg m^{-3}) and MTBE (11.2 μg m^{-3}), and is identified as "gasoline vapor". These VOCs are highly volatile and components of gasoline during the sampling era. The RIOPA samples, collected from 1999 to 2001, reflect the gasoline composition from a decade ago when benzene levels were higher (benzene content is now limited to 0.62% of the fuel (US EPA 2007a). Also, MTBE was used in California, New Jersey, and Texas (US EPA 2008b), but has been phased out (starting in 2000, fully in 2006) (US EPA 2012c).
- Mixture A2 is designated as "vehicle exhaust" due to contributions from toluene (4.9 μg m^{-3}), ethylbenzene (1.9 μg m^{-3}), m,p-xylene (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 as well as exhaust emissions from gasoline- and diesel-powered vehicles (ATSDR 2007, 2010b, a).
- Mixture A3 included 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, e.g., 1,4-DCB is a the major ingredient of mothballs (ATSDR 2006a) (although similar repellents often use naphthalene). PERC is a widely used dry cleaning solvent (ATSDR 1997b). Chloroform is a by-product of water disinfection using chlorine dioxide (ATSDR 1997a). TCE and CTC are used in industry as degreasers, chemical intermediates, and pesticides (ATSDR 1997c, 2005a).
- 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 pinene are widely used flavors and fragrance additives in cleaning products, fresheners, other consumer products, and even in foods and beverages (IARC 1993; US EPA 2012b).

These four mixtures respectively explained 20.5, 20.9, 16.3 and 42.3% of the variation in ΣVOC levels in the RIOPA dataset (Table 38). PMF is often used for source apportions, usually for ambient particulate matter, and these 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 observed in a study using PMF and the NHANES dataset, although NHANES did not measured d-limonene, α-pinene and β-pinene, and the dominant mixtures were gasoline vapor and the vehicle exhaust (Jia et al. 2010). Mixture A4, cleaning products and odorants, explained the largest portion (42.3%) of the total VOC exposure. This large fraction is a result of the VOCs included in RIOPA, the large fraction (87% on average) most people spend indoors (Klepeis et al. 2001), the wide use of the VOCs in this mixture, and their high concentrations (relative to other VOCs measured in RIOPA). Because many of the RIOPA participants were older (average age = 45 years old; 24% were ≥ 60 years old) and predominantly female (75%), we suspected that indoor residential fraction would be especially important. Indoor time fractions calculated for the RIOPA participants, which included indoor at home, school, work, and "other" indoor locations, indicated that RIOPA participants spent an average of 91% of time indoors -- higher than the national data. (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 TVOC.

Identifying the emission source(s) is a key determinant of exposures, and an essential step prior to implementing any exposure reduction strategy. PMF provides a concentration-based approach that can identify generated mixtures, discussed earlier as those that arise from a common or correlated emission source. However, VOC levels also may reflect common contaminant transport and fate factors (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, intentional and possibly coincidental mixtures. It should also be noted that unlike the mixtures based on the mode of actions, the PMF-based mixtures should be orthogonal, that is, uncorrelated.

3.6.5 High Exposure Mixtures

The analysis of high exposure mixtures, which were identified in Section 3.6.4, suggested several variables associated with high exposures (Table 39). When comparing the top quartile to the remainder of the data, the following variables were significant (95 percent confidence interval excluding 1, except as noted):

- City effect: Participants in Los Angeles and Elizabeth had lower odds of high exposure ($\geq 75^{th}$ percentile) than Houston participants for all mixtures (ORs from 0.18 to 0.63), except mixture A3 for the Elizabeth participants.
- Race/ethnicity: Mexicans had increased odds of high exposure to mixtures A1 (benzene and MTBE), A3 (1,4-DCB, TCE, PERC, chloroform and CTC), and A4 (d-limonene, α-pinene and β-pinene) compared to Whites (ORs from 2.03 to 3.97). Hispanics had higher odds of high exposure to mixture A3 than Whites (OR = 1.78, 95% CI = 1.09-2.92). Asians, Blacks and Indians were less likely to have high exposure to mixture A2 (toluene, ethylbenzene, xylene, and styrene) than Whites (OR = 0.47, 95% CI = 0.24-0.92).
- Employment: Employed participants had lower odds of high exposure to mixture A4 (OR = 0.40, 95% CI = 0.27-0.61)
- AERs: Higher log transformed AERs decreased odds of high exposure to all VOC mixtures, especially for mixtures associated with strong indoor sources, e.g., d-limonene and pinene (mixture A4); (ORs from 0.38 to 0.69).
- Open doors or windows: Participants reporting opening doors or windows during the sampling periods had lower odds of high exposure for all mixtures than individuals not opened doors or windows (ORs from 0.32 to 0.40 with 95% CIs not including 1, except for mixture A1). As seen for AERs, this effect of opening doors or windows was more pronounced for mixture A4 (d-limonene and pinene).
- Attached garages: Participants living in houses with attached garages had increased odds of high exposure to mixtures A1 (gasoline vapor) and A2 (vehicle exhaust) mixtures (ORs = 2.27 and 1.95, 95% CIs = 1.45-3.56 and 1.25-3.05, respectively).
- Participant activities: Participants who self-pumped gas during the sampling period had increased odds of high exposure to the gasoline mixture A1 (OR = 2.10, 95% CI =

1.35-3.52). Participants who used fresheners had higher odds of having high exposure to the d-limonene, α-pinene and β-pinene mixture A4 (OR = 2.20, 95% CI = 1.17-4.14).

- Activities of family members: Family members showering during the sampling period had increased odds of high exposures to mixtures A3 (moth repellents, chlorinated solvents and water disinfection by-product mixture, OR = 2.06, 95% CI = 1.20-3.56) and A4 (cleaning and odorant mixtures, OR = 2.45, 95% CI = 1.42-4.23).

Notably, city, ethnicity, and AERs were significantly associated with all VOC mixtures. In addition, several factors identified for gasoline and vehicle exhaust mixtures for the RIOPA participants also have been shown for the personal exposures measurements in NHANES, e.g., the presence of attached garages and self-pumped gas were related to benzene, toluene and MTBE exposures (Jia et al. 2010). However, statistically significant factors have not been identified for 1,4-DCB and chloroform in the NHANES dataset. Factors associated with this mixture may have been identified in RIOPA due to demographic differences between NHANES and RIOPA, specifically, RIOPA participants were more likely to be older, female, unemployed, and at home more often (Su et al. 2012), all of which may increase the importance of indoor sources of 1,4-DCB and chloroform for these participants.

The logistic regression models used do not require normality of the response variables. Thus, even variables with right-skewed distributions do not significantly affect the robustness of the models.

As noted earlier, the main objective of the PMF analysis was to identify mixtures. A more detailed analysis of factors associated with exposure to individual VOCs, that is, the determinants of exposure, and that accounts for repeated measures and interactions, is provided in Section 3.9 using LMMs.

3.6.6 The Robustness of PMF Results

We investigated the robustness of PMF results using the bootstrap method. This method is a re-sampling technique in which "new" datasets are drawn in by randomly selecting observations, and results of the analysis (using PMF) are compared to those obtained using the original data (US EPA 2008a). The variability of the results using the

bootstrap samples shows the stability of original results. We used 500 runs, the original sample size, random sampling with replacement, and the personal VOC exposures. Figure 21 represents the variability for each species of the profiles using box plots. The original results are shown (as a blue box) for reference. Although 2 to 4 of the VOCs in each factor have large variability, e.g., m,p-xylene, MTBE and PERC in the odorant profile, the variability of the VOCs selected to represent the source type in each factor is small, and the original results are consistent with the medians of the bootstrap model results. Thus, source apportionment results using PMF method provided quite robust results.

3.7 Dependency Structures and Joint Distributions of VOC Mixtures in RIOPA

3.7.1 Copulas

The selected copula types are listed in Table 40. (Parameters of the marginal distributions, GOF statistics and copula parameters are in Table 41 to 43.) AICs and BICs for the different copulas were fairly similar for mixtures A1 (benzene, MTBE), A3/B3 (1,4-DCB, TCE, PERC, chloroform, CTC), A4 (d-limonene, α-pinene, β-pinene) and B1 (ethylbenzene, MTBE), however, AICs and BICs for mixtures A2 (toluene, ethylbenzene, xylene, styrene) and B2 (benzene, MTBE, 1,4-DCB, TCE, PERC) were much lower for Gaussian and t copulas, suggesting that these copulas differ in their ability to describe the dependency structures. Gumbel copulas best fitted mixtures A1 and B1, both of which included two VOCs, while t copulas best fitted mixtures A2, A3, A4 and B2, each of which contained four or more 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. Since data below the MDLs were assigned a single value (0.5 MDL), these single values formed "ties" in the distribution. Scatter plots for any two variables that contain many ties display a star shape, which fit the t copula. In contrast, mixtures A1 (benzene and MTBE) and B1 (ethylbenzene and MTBE) contained at least one VOC with very high detection frequencies (e.g., 96% for MTBE), and joint distributions did not show this star shape. Among other mixtures containing at least two VOCs with many non-detects, joint distributions formed star shapes. To explore this explanation, a 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 40 contrasts the probability of exceeding various percentile cut-offs for observed data and that predicted using the copula simulations. Differences were generally small. For the binary mixtures A1 and B1, differences ranged from 0.001 (A1 at the 90th percentile and B1 at 50th, 75th, and 95th percentiles) to 0.02 (B1 at the 75th percentile). For mixtures with three or more components, differences ranged from 0.001 (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 40 also shows crossing probabilities, assuming the mixture components are uncorrelated (independent). As expected, these estimates fell far below observations, especially at higher percentiles, e.g., for the odorant mixture A4 (d-limonene, α-pinene and β-pinene), the observed 90th percentile probability was 0.023, but only 0.001 if the components are assumed to be uncorrelated. Such large differences demonstrate the need to account for dependencies in mixtures.

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

3.7.2 Mixture Fractions

Median mixture fractions are shown in Table 44. The copula simulations matched the mixture fraction for the dominant components observed in all mixtures at all levels, with one exception (mixture B2 at the 75 to 90th percentile level). Often, a single compound dominated the mixture, e.g., MTBE accounted for 78 to 94% of the exposure in mixtures A1 and B1 considering both observations and copula simulations. VOCs with strong indoor sources, e.g., 1,4-DCB and d-limonene, dominated mixtures A3 and A4, respectively, and their fraction increased with percentile. For example, the median fractions of 1,4-DCB in mixture A3 (1,4-DCB, TCE, PERC, chloroform, CTC) for 50-75th percentile observations and simulations

were 0.33 and 0.45, respectively; these increased to 0.99 and 0.99, respectively, at the 95-100th percentile. These results reflect the extreme values previously found for 1,4-DCB and d-limonene (Su et al. 2012). In contrast, mixture fractions varied little for mixtures A1, A2 and B1, e.g., toluene was the dominant component in mixture A2 (toluene, ethylbenzene, xylenes and styrene) with mixture fractions of 0.58 and 0.56 for observations and simulations, respectively, at the 50-75th percentile level, and 0.57 and 0.53, respectively, at the 90-95th percentile. Consistent mixture fractions may suggest generated mixtures as compared to other types where compositions are more varying. Mixture B2 shifted composition at upper percentiles, e.g., the MTBE mixture fractions were 0.61 and 0.55 at the 50-75th percentile levels for observations and simulations, respectively, but 1,4-DCB was dominant at the 95-100th 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. This mixture was selected based on the similar mode-of-action for the component VOCs (and not on the basis of common sources or high correlations). Mixture B2 may be considered an "incidental" mixture as it likely combined 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 Table 41, and mixture fractions in Table 45.) For mixture A3/B3, the analysis revealed only small changes in median fractions, e.g., 1,4-DCB remained the dominant component at high exposure levels, and its mixture fraction increased with percentile. Mixture B2 showed larger differences between median fractions for the (best-fit) t and other copulas, and the dominant VOC at the 90 to 95th percentile level differed among copulas, e.g., the dominant VOCs were 1,4-DCB for the t and Clayton copulas, but MTBE for the Gaussian, Gumbel and Frank copulas. Even though 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.

3.7.3 Estimated Cancer Risks

Estimated cancer risks for the mode-of-action mixtures B1 to B3 are shown in Table 46.

Based on the observed data, VOC mixtures can present rather high cancer risks, e.g., about 10% of RIOPA participants had exposures of mixtures B2 and B3 associated with a 10^{-3} or higher lifetime cancer risk. Mixture B1 (ethylbenzene and MTBE) posed lower risks, e.g., a 25% chance of exceeding a risk of 10^{-5}, and 1% chance of exceeding 10^{-4}. For mixture B2 (benzene, MTBE, 1,4-DCB, TCE and PERC), 3% of participants exceeded a very high risk level, 10^{-2}. Similar results were seen for mixture B3 (1,4-DCB, TCE, PERC, chloroform and CTC).

For each mixture, the copula simulations gave risk predictions that were generally similar to observations, although there is notable divergence at the highest levels, particularly for mixture B3 (Table 46, Figure 22). 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 simulations slightly overestimated the chance of exceeding a risk of 10^{-5}, but underpredicted higher risks. For example, moving vertically on the figure at the risk level of 10^{-5}, the observed data, copula simulations, and lognormal simulations respectively predicted 25, 27 and 32% of individuals in excess of this risk level. At a risk of 10^{-4}, predictions for observed data, copula simulations, and lognormal simulations were 1, 0.6, and 0%, respectively. For mixture B2, lognormal simulations again overestimated low to moderate risks (10^{-6} to 10^{-4}), and 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 22) shows that the copulas sometimes overpredicted the highest values, information not seen in Table 46, e.g., the highest observed risk for mixture B3 was 3.0×10^{-2} while the highest copula simulation was 8.1×10^{-2}. However, such cases were rare (< 1% of the cases).

This analysis suggests that lognormal distributions are a poor choice to represent extreme values, as has been noted earlier (Su et al. 2012). 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 marginal distributions. However, the copula simulations propagate any mismatches in the marginal distributions, which may explain the underprediction of the higher risk levels. Second, copulas permit asymmetric dependency structures that can emphasize

extreme values or other portions of the distribution that display “local” dependencies, e.g., mixture B1 fit the Gumbel copula which emphasizes upper tail dependencies. Lastly, copulas performed better than multivariate lognormal models in all cases, although copulas predictions also diverged from the very highest observations, e.g., above the 95th percentile.

3.7.4 Strengths and Limitations

This is the first study to estimate dependency structures of personal exposures to multivariate VOC mixtures using copulas, a powerful technique that is unrestricted with respect to the marginal distributions of the underlying mixture components. Since VOC exposures were right-skewed even after log-transformation, traditional methods do not properly capture the tail behavior of the VOC distributions. Using the RIOPA data, two sets of VOC mixtures were identified, namely, those based on correlative measures (using PMF analyses), and those based on toxicological mode-of-action. In the former group, the RIOPA data revealed four common mixtures, which were easily identified and considered to be "generated" or "intentional" mixtures. The second group of mixtures, which potentially cause similar health effects, were associated with high lifetime cancer risks, at least for the more exposed individuals. Copulas can improve the precision of exposure estimates, and decrease the bias of risk estimates. Like the cumulative cancer risks predicted in this study, exposures to VOC mixtures should be modeled appropriately to obtain accurate risk estimates. Another application concerns the population attributable fraction (PAF), which quantifies the contribution of various risk factors to a disease, i.e., the number of cases that would not occur if the risk factor did not exist (WHO 2013). In this case, the proportion of population exceeding certain exposure levels, e.g., an exposure threshold, could be estimated to obtain the correct PAF.

The study has several limitations. First, to avoid the effect of repeated measurements, only the first-visit data from RIOPA were used. This decreased the sample size and did not permit the analysis of possible seasonal effects. Second, because PMF does not indicate the optimal number of factors, there is some arbitrariness in this analysis. However, the VOC components in each factor were quite consistent, and the factors often resembled in other studies. The analysis tested only two families of copulas (elliptical and Archimedean) due to the limitations of the software for copula simulations. However, these are best known and

most commonly used copulas. The RIOPA data have some limitations. Only 18 VOCs were measured, and MDLs for some compounds were higher than desirable. Low detection frequencies may affect results of PMF, copula and risk evaluations. While the PMF analysis incorporated uncertainty, distribution and copula selection and fitting assumed that the measurements were error-free. Of course, exposure measurements can involve many types of errors, and both the lowest and highest measurements may be especially prone to errors. The RIOPA sample is not population-based, and results may not be generalizable to the population as a whole. Finally, the RIOPA dataset is over ten years old, and changes in product formulation and other factors may have altered both the concentrations and compositions of VOC exposures.

3.8 Time and VOC Fractions in RIOPA

3.8.1 Time Fractions

Figure 23 displays the average time fractions spent outdoors, indoors and in transit for the RIOPA participants. Indoor time fractions averaged 89, 92, and 92% in Los Angeles, Elizabeth, and Houston, respectively, $p < 0.001$), and participants in Los Angeles spent the least time at home (71, 80, and 80% for the three cities, $p < 0.001$), likely explained in part by the lower unemployment rate in Los Angeles. Little time was spent outdoors, including time within and out of their neighborhoods (fractions averaging 5.1, 4.5, and 4.3% in Los Angeles, Elizabeth, and Houston, respectively, $p = 0.650$). Similarly, time spent in transit was small (5.5, 3.6 and 3.6 in the three cities, respectively, $p < 0.001$).

Figure 23 compares the RIOPA time budgets to a nationally representative sample using the National Human Activity Pattern Survey (NHAPS), a probability-based telephone interview survey conducted from 1992 to 1994 that collected 24-h time-activity information, demographics, and exposure-related questions from 9,196 respondents (Klepeis et al. 2001). NHAPS respondents spent more time outdoors (7.6%) than the RIOPA participants (4.6%), but less time indoors (87%) and at home (69%). This difference may result from the RIOPA's predominating female (75% vs. 54% in NHAPS) rate and older participants (18% of RIOPA participants over 64 years old vs. 14% in NHAPS). Also, the unemployment rate (53%) was high in RIOPA. These older, female and unemployed participants may spend most of their time at home or other indoor places. Indeed, the data from NHAPS shows somewhat more

time in transit and less time at school/work. Both RIOPA and NHAPS reflect the well know pattern that most individuals spend the overwhelming fraction of time at home.

3.8.2 Outdoor and Indoor Exposure Fractions

The home environment dominated personal VOC exposures, e.g., median and mean F_{home} values ranged from 0.63 (MTBE) to 0.78 for α-pinene (Figure 24A, Table 47). The 95th percentile values, which approached to 1 for all VOCs, show an even stronger influence of the home. F_{home} differed by season for two VOCs (benzene and MTBE), and by city for most VOCs (except toluene, o-xylene, 1,4-DCB, PERC, d-limonene and β-pinene). The median F_{home} was highest in Houston (68% to 81%) for most VOCs (except benzene, styrene, PERC, and d-limonene). The importance of the home environment 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 outdoors levels.

Outdoor contributions to personal exposure, shown in Figure 24B, were very small, e.g., median values of $F_{outdoor}$ ranged from 0.02% (d-limonene) to 1% (CTC). Thus, the outdoor environment typically accounted for below 1% of personal exposure, and even less for those VOCs with strong indoor sources, e.g., 1,4-DCB and chloroform. Even the 95th percentile values of $F_{outdoor}$ fell below 15%. $F_{outdoor}$ differed ($p < 0.05$) by season for all VOCs and by city for over half of the VOCs (benzene, toluene, m,p-xylene, o-xylene, MTBE, TCE, PERC and CTC). (Differences by city and season are shown in Table 48.) Outdoor contributions were small, a result of both the little time spent outdoors and the low outdoor VOC concentrations. Because many of VOCs (toluene, styrene, 1,4-DCB, TCE, chloroform, d-limonene, α-pinene, β-pinene) had low detection frequencies (< 60%), the outdoor exposure fractions are approximate.

The two VOC fractions (F_{home} and $F_{outdoor}$) estimated in the study do not represent the whole "exposure profile" contributed by various microenvironments, but this analysis does highlight the most significant contributor of VOC exposures, the home environment. Since this study population mainly comprised older, female, and unemployed participants, who spent most of time at home, the effect of other microenvironments may less important.

The literature is consistent regarding the dominance of the indoor microenvironment for VOC exposure (Lioy et al. 1991; P Kinney et al. 2002; Adgate et al. 2004; Phillips et al. 2005;

Sexton et al. 2007). For example, the home exposures of toluene, styrene, 1,4-DCB, PERC and chloroform dominated exposure for a group of school children (n = 73) in Minneapolis, Minnesota (Adgate et al. 2004). These children spent an average of 65% of their time at home. Time-weighted indoor concentrations were positively associated with personal exposure for these VOC, while time-weighted outdoor concentrations did not have significant associations. In another Minneapolis/St. Paul study, nonsmoking adults (n = 70) showed similar results, with > 50% of VOC exposure occurring at home and 71% of time spent at home (Sexton et al. 2007).

In the present study, indoor VOC levels did not vary seasonally, but city effects were significant, a likely result of differences in emission sources, meteorology and household characteristics (e.g., presence of attached garage) among the three cities studied, as discussed later. Seasonal effects on indoor levels of VOCs in RIOPA may be affected and potentially diminished by lifestyle factors, e.g., opening windows, and using air conditioners. Other important factors affecting indoor concentrations were household characteristics such as the existence of attached garages (Batterman et al. 2007) (also see Section 3.9.5).

3.9 Determinants of Personal, Home, and Outdoor VOC Concentrations in RIOPA

3.9.1 Gasoline-related VOCs

BTEX, MTBE and styrene, all components of gasoline and vehicle exhaust, shared several exposure determinants (Table 49 and Supplemental Table S10). Increased exposures were associated with living in Houston, homes with attached garages, and self-pumped gas; decreased exposures were associated with higher wind speeds and house AERs. Interestingly, lower exposures of toluene, ethylbenzene and o-xylene were found for participants reporting cooking activities during the sampling period, possibly because these individuals drove less for food related activities. Indeed, participants reporting cooking activities spent less time in cars with closed windows (mean time spent = 71 min) than those not reporting 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 (Table 1). In Houston, important VOC sources included petrochemical facilities and vehicles (Weisel et al. 2005b). Attached garages are known sources of gasoline-related aromatics in homes

(Batterman et al. 2007; Sexton et al. 2007; D'Souza et al. 2009; Delgado-Saborit et al. 2009; Symanski et al. 2009; Wang et al. 2009). Gasoline pumping has been shown to elevate personal exposures to BTEX in cold weather in Alaska (Backer et al. 1997). The effects of both attached garages and pumping gas on gasoline-related VOCs were also seen in NHANES (Symanski et al. 2009). Concentrations arising from outdoor sources, e.g., vehicle exhaust, are diluted by wind (US EPA 2010b), so higher wind speeds may lower exposures. The AER, which accounts for infiltration and ventilation and which depends on wind speed (US EPA 2011b), influences indoor concentrations and thus personal exposures for those pollutants arising from indoor sources. Cooking-related activities have been shown to increase indoor and personal concentrations of several VOCs, e.g., benzene and toluene (Clobes et al. 1992; Byun et al. 2010). However, in RIOPA, 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 does not allow further analysis, but we speculate that visits to "drive-though" fast-food facilities where vehicles are queued up and idling may be a particularly important source of VOC exposure.

3.9.2 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 50 and Supplemental Table S11). Like the gasoline-related VOCs, Houston participants had higher exposures to these VOCs. AERs were negatively associated with VOC exposures, reflecting the dilution effects affecting indoor sources. Participants in larger houses (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, e.g., 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, 2006a; Chin et al. 2012; US 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 house size, as a determinant may reflect additional mixing in large houses that lowers concentrations compared to approximately the same product use in smaller houses. We have 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 RIOPA, the number of rooms in a household was positively associated with household income (β = 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, no association with household income and VOC exposures were found. The effect of employment on d-limonene exposure might result as unemployed participants spent more time at home (2,278 and 2,000 min 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 study 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 in RIOPA increased when other 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 dataset, 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 since 87% of participants showered during the sampling period. Additionally, participants were instructed not to get the samplers wet, and they may have removed them outside of the shower and 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 (Weisel et al. 2005b), meteorological factors that affect both dispersion and emissions of outdoor pollutants, systematic differences in building

AERs, demographic and cultural factors. For example, outdoor temperatures were considerably warmer in Houston during the sampling period, compared to Los Angeles and Elizabeth (respectively averaging 22.3 ± 7.5, 18.6 ± 4.7 and 14.6 ± 8.6 °C, p-value < 0.001). Higher temperatures increase vapor pressures, permeation rates, and evaporation rates, potentially producing higher concentrations. Since a fraction of odorant and cleaning-related VOCs arise from volatilization and sublimation from indoor sources, indoor temperatures are also important. Indoor temperatures showed less variation and differences were not significant (respectively averaging 23.3 ± 2.6, 23.9 ± 2.6°C and 24.0 ± 3.4 in Los Angeles, Elizabeth, and Houston, p-value = 0.052).

3.9.3 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 51 and Supplemental Table S12). 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 (Table 51 and Supplemental Table S12). This solvent has been widely used for dry cleaning clothes, and exposures occur when visiting dry cleaning establishments, and storing dry cleaned clothes at home, whether or not clothes are wrapped in plastic (Sherlach et al. 2011), as noted in Table 1. PERC exposures were higher among employed participants. Since PERC has been widely used in industry as a degreaser and also has been added into products such as adhesives and paint removers (ATSDR 1997b), employed participants may have more chances to contact it. The city effect may be related to population density: Los Angeles and Elizabeth have higher densities (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 were 1.29, 0.74, and 0.11 μg m^{-3}, respectively, p-value < 0.001).

TCE has been used extensively as a degreaser, paint remover, adhesive, and chemical intermediate (ATSDR 1997c). Exposure may increase if TCE-containing consumer or home products are present, e.g., vinyl siding, glue and car stain removers (US EPA 2007b). Additionally, TCE is sometimes found in contaminated soils and groundwater, and

participants in households near to subsurface or surface contaminated soils may be exposed indoors through soil vapor intrusion and water consumption, if a local well (especially a private well without water monitoring or treatment) provides the water source. In the RIOPA dataset, the TCE detection frequency was only 31%, thus, the 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 2005a). (Previously, CTC 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 are no effective substitutes. CTC is globally distributed at generally low levels with spatial little variation, except near contaminated source areas where levels increase. The variation among CTC exposures among the RIOPA participants is limited, and little variance can be explained by the available variables.

3.9.4 Summary of Key Exposure Determinants

The most common and significant determinants of personal VOC exposures were city, inverse wind speed, log-transformed AER, number of rooms, presence of an attached garage, and self-pumping gas. Inverse wind speed was positively associated with log-transformed benzene, ethylbenzene, m,p-xylene, o-xylene, MTBE, and PERC. Log-transformed AER was negatively associated with log-transformed toluene, ethylbenzene, m,p-xylene, o-xylene, PERC, chloroform, d-limonene, α-pinene and β-pinene. Participants living in larger houses (more rooms) had lower exposures of benzene, styrene, 1,4-DCB, chloroform, d-limonene, and α-pinene; those in houses with attached garages had higher levels of benzene, toluene, ethylbenzene, m,p-xylene, o-xylene, and MTBE. Participants who self-pumped gas had higher exposures of benzene, ethylbenzene, m,p-xylene, o-xylene, and MTBE. While 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 50 and 51). These effects were significant and based on linear mixed models, which controlled for clustering and repeated measures. As discussed later, the LMMs explained for 0.003 (CTC) to 0.4 (β-pinene) of the variance in personal

exposure.

3.9.5 Determinants of Indoor VOC Concentrations

An analysis parallel to that performed for personal samples, i.e., using LMMs, was conducted for the indoor VOC measurements. Given the correlation between indoor and personal exposure measurements, it is not surprising that many of the same factors were identified as determinants (Tables 52 to 54). 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-xylene, o-xylene, styrene, TCE, PERC, chloroform, d-limonene, α-pinene and β-pinene. Larger houses (more rooms) was associated with decreased concentrations of benzene, toluene, m,p-xylene, 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. (These were highest in Elizabeth).

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

3.9.6 Determinants of Outdoor VOC Concentrations

Outdoor concentrations were affected by city and three meteorological variables (Tables 55 to 57). Ambient relative humidity was negatively associated with concentrations of benzene, ethylbenzene, m,p-xylene, o-xylene, MTBE, styrene, and β-pinene levels. Wind speed was negatively associated with concentrations of benzene, toluene, ethylbenzene, m,p-xylene, o-xylene, MTBE, styrene, TCE, PERC, and α-pinene. Effects of city and

outdoor temperature depended on the VOC. For example, Houston had the highest concentrations for benzene, m,p-xylene and β-pinene, which may be due to the crowed petrochemical industry (Weisel et al. 2005b).

3.9.7 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. That five common factors affected concentrations of most personal and indoor VOC measurements suggests that the critical influence of indoor sources (or levels) on personal exposures. In contrast, outdoor levels had only minor impacts on personal exposure, although they may influence indoor levels (Sexton et al. 2007). As in many other studies, RIOPA participants spent most of their time indoors, and outdoor concentrations were low.

3.9.8 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 account for effects of all other covariates. Plots for wind speed and AER suggested some non-linearities with log-transformed VOC concentrations (Figures 25A, C, and E). Several transformations of these variables were attempted, and near-linear relationships were achieved using the reciprocal of wind speed and the logarithm of AER (Figures 25B, D, and F). Inverse wind speed can be supported based on dilution or mass balance principles (applying 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 is expected 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 AER may have had some outliers and possibly some bias or errors. Still, the expected relationship was seen, i.e., indoor concentrations of VOCs with strong indoor sources (e.g., chloroform and d-limonene) decreased as AERs increased (Table 53).

3.9.9 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 Table 58. For personal exposures, R^2 ranged from 0.003 (CTC) to 0.40 (β-pinene); for indoor measurements, the R^2 ranged from 0.09 (toluene) to 0.42 (PERC); and for outdoor concentrations, the R^2 values were from 0.17 (1,4-DCB) to 0.65 (PERC). Generally, more variance was explained for the outdoor measurements. VOCs with specific emission sources, e.g., PERC (dry cleaners) and α-pinene (cleaning products and freshener), had the largest R^2 among 15 VOCs; this applied to all three sample types. In contrast, VOCs used in many commercial products and that were also components of exhaust and 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 dataset. While some of the variance is random and some is due to errors in measurement and model specifications, 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. However, low R^2 values do not invalidate the identification or significance of the determinants.

3.9.10 Strengths and Limitations

The analysis of the extended and comprehensive RIOPA dataset, which includes outdoor, indoor and personal measurements of 15 VOCs along with over 500 other variables used as candidate factors, advances the understanding of VOC exposure and exposure determinants. The relationship of outdoor and home VOC levels to personal exposures were evaluated, using time and VOC fractions, and many factors were shared among outdoor, indoor and personal measurements. Strengths of analysis include the use of LMMs, the repeated measurements for available participants, and the nested analysis, which allowed estimation of individual differences from average levels for specific variables (Krueger and Tian, 2004; Wu, 1996). The time fractions help to understand the participants' activity pattern, and to estimate the contribution of VOC sources to exposures. Many of our results are consistent with previous studies, e.g., the significance of strong indoor VOC sources (Sexton et al., 2007), the presence of attached garages (D'Souza et al., 2009; Delgado-Saborit et al., 2009; Sexton et al., 2007; Symanski et al., 2009; Wang et al., 2009), and activities such as visiting dry cleaners (D'Souza

et al., 2009; Wallace, 1989; Wallace, 2001; Wang et al., 2009). Several new determinants were discovered, including a strong effect due to city, other family member showering, and residence size.

The limitations of the dataset include missing data, which decrease sample size and statistical power. Two methods were used to address this issue. First, variables with sample sizes less than 400 (>150 missing cases) were excluded from LMMs. This excluded several potentially significant variables, e.g., land use data. Fortunately, land use data were highly correlated with city, which was utilized in every model. Second, the use of multiple imputations was evaluated, and results showed that for the models tested, the impacts of missing data would not be substantial. We also noted that models for personal exposures explained less variance (lower R^2) than outdoor and indoor models, probably due to the number and complexity of factors (especially behaviors) that affect an individual's exposure. A final limitation of the study is the representativeness of the study sample. RIOPA data was collected in three U.S. cities, which have specific emission sources (Weisel et al. 2005b). A convenience sample was used, which led to a number of demographic and other differences, as discussed. Since the study period, VOC sources and levels may have changed somewhat. Thus, study results may not reflect the U.S. population or current period. However, most findings correspond to other studies that using regional or national data, thus, most of the results appear relevant.

Table 1. Determinants of VOC exposures in previous and present studies.

Determinants	Benzene	Toluene	Ethylbenzene	m,p-Xylene	o-Xylene	MTBE	Styrene	1,4-DCB	TCE	PERC	Chloroform	CTC	d-Limonene	α-Pinene	β-Pinene
Personal activities															
Contact with chlorinated water								m			A, C, M		M	M	M
Cooking	L	L, m	m		m						L				
Cycling/ walking		E	E	E	E										
Keep pets											m				m
Near vehicle or engines	D, E, G	D	A, D, E	A, D, E	A, D, E		D			A					
Polish/wax furniture				j	j			M							
Pump gas/near gasoline	E, K, M	J, K	E, J, K, M	E, J, K, M	E, J, K, M	M									
Renovate house		M											M		
Smoke or near ETS	A, B, C, D, G, H, k	B, D, e, H	B, D, H	B, D, H	B, D, H		A, B, D								
Stay in/ presence of attached garages	F, G, H, J, K, M	F, H, J, M	F, G, H, J, K, M	F, H, J, K, M	F, H, J, K, M	H, M				H					
Time spent at home		m					m								
Time spent in closed cars									M						
Undertake arts and crafts		E	E	E	E										
Use air cleaning devices				M	M							M			
Use deodorizers and mothballs								A, C, H				m			
Use gas heating/gas stove	D, G, M	D, j	D	D	D	D	D						M		
Use paint and other solvents	H	H, K	G, H, K, M	H, J, K, M	H, J, K						K				
Use perfume						m									
Visit dry-cleaner/near dry-cleaned clothes										A, C, H, K, M					
Socioeconomic factors															
Age											i, k				
City/ region*	l, m	l	l, m	l, m		m	m	m	m	m	m	m	m	m	m
Education/parental education	k				l			k							
Non-Hispanic White	h, k	h	h	h	h	h		h, k			h, i, k				
Male	K		K	K	K						k				
Machine-related jobs/ work in a factory	H	H	G, H	H	H										
Ownership of the house											m				
Unemployed										m			M		
Environmental factors															
AER		m	m	m	m					m	m		m	m	m
Ambient RH										m	m				m
Furniture refinisher in neighborhood								M							
Existence of a fireplace							G					M			
Existence of a swim pool											H, I			M	
Existence of a well/ use well water									M		h	m			
Indoor temperature	m								m						
Live in an apartment/mobile home	L										I				
Near commerical street/ highway						H		H		H					
Number of floors	m					m									
Number of rooms	m						m	m			m		m	m	
Open windows/ doors	f, h, j, k	f, h, j, k	f, h	f, h	f, h, m	f	f, m	f, m		f, h	f, h, i, k		f	f	f
Restaurants or bakery in neighborhood								M	m						
Vinyl, asbestos or other siding									M						
Wind speed	m		m	m	m	m				m					
Years lived in home	h	h	h	h	h										

A, Wallace et al. 1989; b, Edwards et al. 2001; c, Wallace 2001; d, Kim et al. 2002; e, Hinwood et al. 2007; f, Sexton et al. 2007; g, Delgado-Saborit et al. 2009; h, D'Souza et al. 2009; i, Riederer et al. 2009; j, Symanski et al. 2009; k, Wang et al. 2009; l, Byun et al. 2010; m, the present study.
Capital letters indicate increased exposure, and lower case indicates decreased exposure; *, no increasing or decreasing trends.

Table 2. Toxicity standards/guidelines for personal VOC exposures.

VOCs	Cancer					Non-cancer			
	IRIS	IARC	URF $(\mu g\ m^{-3})^{-1}$	Source	Health effect	Chronic RfC $(\mu g\ m^{-3})$	Source	Acute MRL $(\mu g\ m^{-3})$	Source
Benzene	A	1	7.8×10^{-6}	IRIS, 2011	Leukemia (occupational)	30	IRIS, 2011	29	ATSDR, 2010
Toluene	D	3	NA	IRIS, 2011	Neurological effects (occupational); color vision impairment (occupational) and respiratory irritation (human volunteer)	5000	IRIS, 2011	3766	ATSDR, 2010
Ethylbenzene	D	2B	2.5×10^{-6}	OEHHA, 2005	Lung, liver, and renal adenomas and carcinomas (animal)	1000	IRIS, 2011	21696	ATSDR, 2010
Xylenes	D	3	NA	IRIS, 2011	Impaired motor coordination (animal)	100	IRIS, 2011	8679	ATSDR, 2010
MTBE	D	3	2.6×10^{-7}	OEHHA, 2005	Lymphomas, leukaemias, hepatocellular adenomas, and renal tubular and testicular tumours (animal)	3000	IRIS, 2011	7206	ATSDR, 2010
Styrene	ND	2B	2.0×10^{-6}	Caldwell et al., 1998	Pulmonary adenomas (animal)	1000	IRIS, 2011	21286	ATSDR, 2010
1,4-DCB	ND	2B	1.1×10^{-5}	OEHHA, 2005	Liver and kidney tumor, and mononuclear-cell leukemia (animal)	800	IRIS, 2011	12019	ATSDR, 2010
TCE	ND	2A	2.0×10^{-6}	OEHHA, 2005	Liver and biliary tract cancer, and lymphoma (human); liver, renal-cell, lung and testicular tumours, and lymphomas (animal)	40	EPA, 2001	10741	ATSDR, 2010
PERC	ND	2A	5.9×10^{-6}	OEHHA, 2005	Oesophageal and cervical cancer, and non-Hodgkin's lymphoma (human); hepatocellular carcinomas and mononuclear-cell leukaemia (animal)	16	EPA, 2010	1356	ATSDR, 2010
Chloroform	B2	2B	2.3×10^{-5}	IRIS, 2011	Renal tubule and hepatocellular tumours (animal)	NA		488	ATSDR, 2010
CTC	B2	2B	1.5×10^{-5}	IRIS, 2011	Liver and mammary neoplasms (animal)	100	IRIS, 2011	NA	
d-Limonene	ND	ND	NA			NA		NA	
α-Pinene	ND	ND	NA			NA		NA	
β-Pinene	ND	ND	NA			NA		NA	

IRIS, Integrated Risk Information System; IARC, International Agency for Research on Cancer; URF, unit risk factor; RfC, reference concentration; MRL, minimal risk level; NA, not available; ND, no data.

Table 3. Statistics of outdoor VOC (μg m^{-3}) concentrations in RIOPA.

Outdoor	n	Mean	SD	GM	GSD	Min	25th	50th	75th	95th	Max
Benzene	555	2.15	2.11	1.57	2.19	0.27	0.69	1.68	2.67	5.16	20.92
CTC	555	0.72	1.31	0.63	1.50	0.14	0.55	0.64	0.75	1.00	31.23
Chloroform	555	0.37	1.43	0.22	1.87	0.14	0.14	0.21	0.21	0.79	24.72
1,4-DCB	555	2.15	17.16	0.57	2.69	0.22	0.46	0.46	0.64	3.66	355.05
Ethylbenzene	555	1.28	1.87	0.88	2.29	0.11	0.37	0.93	1.67	3.04	36.24
d-Limonene	555	1.97	6.34	0.78	2.65	0.35	0.35	0.64	0.64	6.54	74.20
MC	555	1.06	2.23	0.63	2.73	0.15	0.15	1.05	1.05	2.46	39.86
MTBE	555	8.11	9.99	5.04	2.79	0.19	2.84	5.32	9.72	22.09	105.17
α-Pinene	555	1.31	4.16	0.71	2.53	0.14	0.46	1.02	1.02	2.23	63.17
β-Pinene	555	0.94	2.15	0.72	1.69	0.51	0.51	0.51	1.05	1.26	46.17
Styrene	555	0.58	2.06	0.39	1.94	0.17	0.17	0.42	0.42	1.29	47.00
Toluene	555	6.83	6.54	5.26	1.91	3.35	3.35	3.56	8.71	19.63	64.97
TCE	555	0.34	1.30	0.22	1.92	0.12	0.12	0.22	0.22	0.80	30.07
PERC	555	1.02	2.17	0.51	3.16	0.11	0.21	0.61	1.21	3.17	41.82
m,p-Xylene	555	3.56	4.16	2.44	2.36	0.33	1.49	2.49	4.26	10.02	51.21
o-Xylene	555	1.46	3.90	0.92	2.31	0.15	0.43	0.96	1.58	3.23	80.98

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; min. minimum; max, maximum.

Table 4. Statistics of outdoor VOC concentrations (μg m^{-3}) stratified by city in RIOPA.

Outdoor	Los Angeles, CA (n=175)						Elizabeth, NJ (n=182)						Houston, TX (n=198)					
	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95th
Benzene	2.50	2.37	1.76	2.36	1.98	6.10	1.45	1.56	1.09	2.03	1.22	3.30	2.48	2.17	1.98	1.93	1.94	5.69
CTC	0.68	0.23	0.64	1.45	0.63	1.00	0.84	2.28	0.63	1.78	0.69	1.04	0.63	0.15	0.62	1.21	0.62	0.80
Chloroform	0.40	0.77	0.28	1.88	0.21	1.26	0.47	1.93	0.25	1.87	0.21	0.97	0.26	1.34	0.16	1.56	0.14	0.35
1,4-DCB	1.32	2.13	0.78	2.34	0.46	5.05	3.58	26.97	0.64	2.63	0.46	6.95	1.57	12.38	0.38	2.71	0.22	2.45
Ethylbenzene	1.61	1.53	1.15	2.30	1.30	4.50	1.34	2.75	0.86	2.31	0.99	2.93	0.94	0.80	0.72	2.12	0.79	2.49
d-Limonene	3.33	9.17	1.30	2.95	0.64	12.30	1.99	5.65	0.88	2.34	0.64	10.90	0.74	2.53	0.44	1.86	0.35	1.36
MC	1.59	3.18	1.22	1.62	1.05	3.25	1.46	2.07	1.19	1.57	1.05	3.09	0.23	0.16	0.19	1.68	0.15	0.59
MTBE	10.79	11.43	7.26	2.61	8.31	26.81	5.77	5.34	3.77	2.75	4.32	19.16	7.89	11.31	4.78	2.72	4.52	25.67
α-Pinene	2.30	6.52	1.27	2.01	1.02	6.52	1.34	3.10	1.09	1.43	1.02	1.02	0.41	0.71	0.29	2.13	0.30	0.84
β-Pinene	0.86	1.43	0.62	1.79	0.51	2.22	0.89	3.46	0.56	1.64	0.51	1.23	1.08	0.46	1.06	1.15	1.05	1.05
Styrene	0.71	0.94	0.53	1.81	0.42	2.52	0.72	3.46	0.45	1.52	0.42	0.82	0.34	0.36	0.25	1.98	0.17	1.09
Toluene	8.69	8.82	6.32	2.10	3.35	24.14	6.80	5.68	5.29	1.93	3.35	18.06	5.21	4.04	4.45	1.62	3.56	14.36
TCE	0.29	0.30	0.25	1.52	0.22	0.59	0.60	2.22	0.36	1.98	0.22	1.05	0.14	0.09	0.13	1.36	0.12	0.30
PERC	1.85	1.90	1.28	2.43	1.30	4.40	1.10	3.09	0.72	2.12	0.74	2.19	0.22	0.20	0.17	1.89	0.11	0.69
m,p-Xylene	4.91	5.25	3.19	2.62	3.56	12.97	3.21	4.31	2.25	2.24	2.34	8.75	2.69	2.18	2.07	2.10	2.23	7.52
o-Xylene	1.78	1.66	1.26	2.31	1.40	4.45	1.67	6.54	0.88	2.17	0.94	2.61	0.99	0.84	0.73	2.25	0.80	2.45

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation.

Table 5. Statistics of indoor VOC (μg m^{-3}) concentrations in RIOPA.

Indoor	n	Mean	SD	GM	GSD	Min	25th	50th	75th	95th	Max
Benzene	554	3.50	5.15	2.19	2.50	0.55	1.27	2.19	3.85	10.03	46.07
CTC	554	0.71	0.97	0.61	1.61	0.14	0.52	0.62	0.75	1.10	18.07
Chloroform	554	1.86	2.97	0.93	3.20	0.14	0.37	0.92	2.16	6.34	40.18
1,4-DCB	554	68.84	303.76	2.61	8.94	0.22	0.46	1.40	7.85	343.88	4050.73
Ethylbenzene	554	2.52	4.74	1.49	2.52	0.29	0.89	1.46	2.47	7.62	68.37
d-Limonene	554	30.98	107.06	9.27	4.81	0.35	3.16	9.67	27.99	102.75	2101.31
MC	554	2.40	10.61	0.91	3.00	0.15	0.67	1.05	1.05	7.50	187.64
MTBE	554	11.79	27.29	5.60	3.26	0.19	3.10	5.98	10.68	36.00	348.04
α-Pinene	554	7.04	14.60	3.03	3.24	0.40	1.02	2.60	7.17	25.49	174.67
β-Pinene	554	4.85	10.95	1.77	3.63	0.51	0.51	1.21	4.46	20.45	123.14
Styrene	554	1.47	4.24	0.68	2.58	0.17	0.42	0.42	1.07	5.13	59.37
Toluene	554	15.26	24.48	9.83	2.37	3.35	3.56	10.41	17.10	39.79	323.95
TCE	554	0.97	7.19	0.27	2.58	0.12	0.12	0.22	0.28	1.73	132.32
PERC	554	1.84	4.47	0.80	3.40	0.11	0.35	0.84	1.71	6.01	78.05
m,p-Xylene	554	7.32	15.87	4.01	2.68	0.33	2.28	4.07	6.91	22.18	231.22
o-Xylene	554	2.47	4.78	1.46	2.51	0.15	0.88	1.46	2.44	7.24	66.88

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; min. minimum; max, maximum.

Table 6. Statistics of indoor VOC concentrations (μg m^{-3}) stratified by city in RIOPA.

Indoor	Los Angeles, CA (n=174)						Elizabeth, NJ (n=182)						Houston, TX (n=198)					
	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95th
Benzene	3.00	5.00	1.94	2.36	2.05	6.53	2.51	3.97	1.53	2.49	1.65	7.33	4.85	5.93	3.38	2.19	3.06	12.23
CTC	0.80	1.69	0.60	1.73	0.58	1.09	0.66	0.30	0.58	1.73	0.63	1.18	0.68	0.28	0.65	1.35	0.62	1.12
Chloroform	1.57	2.13	0.88	2.96	0.92	5.16	1.65	3.46	0.74	3.25	0.74	6.51	2.31	3.07	1.22	3.21	1.32	9.11
1,4-DCB	38.81	315.6	1.61	5.16	1.18	31.06	29.20	121.1	2.40	6.66	1.39	137.1	131.7	389.6	4.32	15.01	2.02	1017
Ethylbenzene	2.45	3.51	1.51	2.53	1.45	7.99	2.30	5.71	1.21	2.69	1.29	7.02	2.78	4.72	1.77	2.26	1.68	7.62
d-Limonene	21.87	45.00	6.96	4.76	7.31	92.25	14.66	24.53	5.41	4.46	6.71	62.56	53.99	170.3	19.58	3.87	20.79	166.8
MC	1.86	2.77	1.33	1.85	1.05	6.67	1.77	4.17	1.24	1.73	1.05	3.71	3.44	17.08	0.49	4.36	0.37	11.80
MTBE	13.16	33.09	6.38	3.31	7.44	26.92	7.35	9.56	3.98	3.31	4.96	25.02	14.67	31.88	6.84	2.97	5.82	55.08
α-Pinene	6.82	14.62	2.57	3.35	1.02	32.60	3.97	10.83	1.92	2.58	1.02	14.34	10.06	16.88	5.35	3.01	5.53	34.90
β-Pinene	3.04	9.20	1.14	3.09	0.51	10.50	3.32	11.15	1.07	3.14	0.51	9.90	7.84	11.57	4.13	3.02	4.03	24.96
Styrene	1.30	2.04	0.71	2.49	0.42	6.45	1.50	4.05	0.64	2.52	0.42	6.60	1.58	5.64	0.68	2.71	0.67	3.04
Toluene	16.29	33.73	9.72	2.45	10.71	34.60	12.75	11.58	9.31	2.19	9.74	34.80	16.66	23.47	10.42	2.46	10.51	47.65
TCE	0.51	2.52	0.26	1.78	0.22	0.62	0.97	2.50	0.47	2.62	0.22	2.79	1.38	11.55	0.16	2.50	0.12	0.85
PERC	3.32	7.06	1.71	2.83	1.66	13.80	1.32	1.98	0.94	2.14	0.90	3.38	1.02	2.42	0.35	3.43	0.30	5.14
m,p-Xylene	6.88	9.33	4.01	2.82	4.16	25.22	6.50	18.99	3.23	2.71	3.18	15.85	8.47	17.23	4.90	2.42	4.55	25.02
o-Xylene	2.44	3.14	1.58	2.44	1.64	7.71	2.13	5.32	1.19	2.52	1.18	6.38	2.80	5.40	1.64	2.49	1.53	8.98

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation.

Table 7. Statistics of personal adult VOC (μg m^{-3}) concentrations in RIOPA.

Adult	n	Mean	SD	GM	GSD	Min	25^{th}	50^{th}	75^{th}	95^{th}	Max
Benzene	544	3.64	5.31	2.37	2.41	0.27	1.45	2.39	4.09	10.74	85.37
CTC	544	0.80	2.44	0.61	1.66	0.14	0.53	0.62	0.74	1.08	42.27
Chloroform	544	4.25	52.49	1.05	3.17	0.14	0.49	1.04	2.20	6.58	1223.56
1,4-DCB	544	56.83	229.37	2.98	8.10	0.22	0.46	1.88	8.30	314.50	2153.45
Ethylbenzene	544	2.78	5.13	1.65	2.54	0.11	0.97	1.68	2.69	7.48	64.55
d-Limonene	544	41.14	238.90	10.90	4.58	0.35	4.85	11.77	29.42	112.21	5113.77
MC	544	3.11	17.14	0.99	3.05	0.15	0.93	1.05	1.05	7.40	329.85
MTBE	544	14.77	42.67	6.98	3.23	0.19	3.83	7.14	13.99	42.67	843.74
α-Pinene	543	6.86	16.26	3.20	3.03	0.55	1.02	2.88	6.95	23.62	231.48
β-Pinene	544	5.53	13.07	1.92	3.72	0.51	0.51	1.52	4.49	22.43	133.16
Styrene	544	1.55	4.31	0.73	2.56	0.17	0.42	0.42	1.10	5.52	59.52
Toluene	544	19.12	37.31	11.60	2.48	3.35	4.38	12.42	19.94	50.25	641.47
TCE	544	1.44	10.74	0.29	2.90	0.12	0.12	0.22	0.47	2.38	200.31
PERC	544	7.17	112.35	0.94	3.54	0.11	0.41	0.89	2.00	7.24	2617.79
m,p-Xylene	544	8.07	15.49	4.63	2.63	0.70	2.71	4.42	7.85	22.73	219.05
o-Xylene	544	2.87	5.59	1.74	2.42	0.42	1.06	1.72	2.77	8.16	79.56

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; min. minimum; max, maximum.

Table 8. Statistics of personal adult VOC concentrations (μg m^{-3}) stratified by city in RIOPA.

Adult	Los Angeles, CA (n=174)						Elizabeth, NJ (n=171)						Houston, TX (n=199)					
	Mean	SD	GM	GSD	50th	95th	Mean	SD	GM	GSD	50th	95th	Mean	SD	GM	GSD	50th	95th
Benzene	3.10	6.53	2.08	2.25	2.26	6.22	2.80	4.19	1.70	2.51	1.76	10.09	4.82	4.76	3.55	2.10	3.13	14.78
CTC	0.86	3.16	0.60	1.66	0.58	0.99	0.88	2.96	0.57	1.99	0.64	1.23	0.67	0.23	0.65	1.28	0.62	1.06
Chloroform	8.52	92.67	0.92	3.14	0.87	5.19	2.20	4.83	0.95	3.38	0.85	7.02	2.27	2.95	1.28	2.95	1.33	8.65
1,4-DCB	14.95	63.36	1.70	4.74	1.23	60.10	26.49	113.7	2.59	6.04	1.85	86.76	119.5	351.3	5.49	12.66	3.42	945.9
Ethylbenzene	2.33	3.60	1.50	2.46	1.66	5.55	2.91	6.89	1.42	2.86	1.40	8.04	3.06	4.45	2.03	2.25	1.83	11.22
d-Limonene	48.17	388.2	7.41	4.43	7.62	87.71	17.91	32.12	6.72	4.58	8.39	59.67	54.97	152.6	23.16	3.43	22.40	154.3
MC	3.84	25.02	1.41	2.15	1.05	8.40	1.81	3.79	1.23	1.78	1.05	4.25	3.60	15.63	0.59	4.39	0.45	12.46
MTBE	12.23	13.48	8.09	2.69	8.52	35.20	14.63	65.43	5.06	3.61	5.49	38.02	17.12	33.88	8.08	3.21	7.32	66.77
α-Pinene	4.83	8.12	2.33	2.92	1.02	26.59	5.06	16.44	2.25	2.73	1.02	15.89	10.17	20.44	5.72	2.68	5.83	27.49
β-Pinene	2.80	8.55	1.06	2.95	0.51	10.09	5.14	16.19	1.32	3.61	0.51	30.34	8.25	12.86	4.45	2.94	4.16	25.53
Styrene	1.19	1.90	0.67	2.39	0.42	6.07	1.84	5.56	0.69	2.76	0.42	10.40	1.61	4.59	0.81	2.51	0.84	3.09
Toluene	18.79	49.32	11.17	2.44	12.71	48.60	20.74	38.63	11.33	2.67	11.33	56.71	18.01	20.17	12.25	2.36	13.09	49.91
TCE	0.72	3.31	0.30	2.20	0.22	1.56	2.39	15.63	0.53	3.03	0.50	4.80	1.26	9.79	0.17	2.62	0.12	1.22
PERC	3.79	9.59	1.86	2.69	1.75	9.82	17.36	200.1	1.11	2.98	1.00	4.94	1.38	4.75	0.44	3.41	0.36	6.40
m,p-Xylene	7.07	9.76	4.45	2.64	4.54	18.89	7.91	20.34	3.84	2.89	4.04	25.51	9.07	14.75	5.64	2.32	5.10	32.12
o-Xylene	2.53	3.42	1.76	2.29	1.84	6.01	3.04	8.06	1.48	2.71	1.56	9.16	3.02	4.40	1.98	2.24	1.80	9.52

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation.

Table 9. Statistics of personal child VOC (μg m^{-3}) concentrations in RIOPA.

Child	n	Mean	SD	GM	GSD	Min	25th	50th	75th	95th	Max
Benzene	209	4.16	5.57	2.84	2.29	0.55	1.75	2.79	4.55	11.95	54.68
CTC	209	0.57	0.16	0.54	1.41	0.14	0.47	0.56	0.67	0.83	1.22
Chloroform	209	2.03	3.63	1.10	2.82	0.14	0.53	1.14	2.12	7.47	38.59
1,4-DCB	209	121.56	313.58	6.83	11.39	0.22	1.05	4.18	25.88	978.59	1783.50
Ethylbenzene	209	3.34	6.35	2.00	2.44	0.11	1.22	1.95	3.07	10.28	60.24
d-Limonene	209	32.11	49.75	16.48	3.56	0.64	8.02	17.36	37.99	111.49	577.74
MC	209	1.70	6.50	0.62	3.31	0.15	0.15	0.88	1.05	5.25	88.88
MTBE	209	11.69	22.06	6.73	2.87	0.19	3.86	7.03	13.46	30.16	224.83
α-Pinene	209	5.69	5.75	3.63	2.63	0.75	1.50	3.57	8.14	16.61	36.03
β-Pinene	209	5.33	6.21	2.79	3.29	0.51	1.05	2.85	8.06	18.22	35.29
Styrene	209	1.70	4.36	0.78	2.65	0.17	0.42	0.65	1.23	6.89	39.70
Toluene	209	18.30	27.82	11.72	2.38	3.35	7.64	12.34	19.49	57.17	238.39
TCE	209	0.35	0.89	0.20	2.17	0.12	0.12	0.12	0.22	0.95	9.62
PERC	209	2.82	15.91	0.67	3.52	0.11	0.29	0.57	1.40	7.34	211.10
m,p-Xylene	209	8.87	16.74	5.31	2.47	0.70	3.14	5.15	8.55	28.17	205.41
o-Xylene	209	2.91	4.88	1.89	2.33	0.15	1.22	1.96	2.89	7.97	59.65

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; min. minimum; max, maximum.

Table 10. Statistics of personal child VOC concentrations (μg m^{-3}) stratified by city in RIOPA.

Child	Los Angeles, CA (n=33)						Elizabeth, NJ (n=41)						Houston, TX (n=135)					
	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95^{th}	Mean	SD	GM	GSD	50^{th}	95th
Benzene	2.20	1.39	1.76	2.07	2.21	5.48	2.75	4.00	1.80	2.37	1.97	6.54	5.07	6.37	3.68	2.08	3.96	13.64
CTC	0.49	0.17	0.46	1.57	0.51	0.77	0.62	0.22	0.57	1.62	0.63	0.94	0.57	0.12	0.56	1.25	0.56	0.80
Chloroform	0.93	0.97	0.61	2.51	0.71	3.78	2.62	6.16	0.99	3.57	1.06	9.07	2.12	2.92	1.32	2.53	1.41	11.40
1,4-DCB	1.33	1.91	0.79	2.45	0.46	6.51	36.40	159.3	2.75	6.76	1.46	137.1	176.8	369.1	15.26	11.23	10.19	1086
Ethylbenzene	1.87	1.14	1.51	2.04	1.69	4.41	3.41	9.24	1.56	2.80	1.53	6.61	3.67	6.01	2.31	2.37	2.08	11.17
d-Limonene	9.96	8.45	6.25	3.17	7.55	31.81	19.86	23.56	9.03	4.68	14.49	52.51	41.24	58.39	25.07	2.66	24.46	126.1
MC	1.44	1.33	1.22	1.57	1.05	5.16	2.22	3.25	1.44	2.11	1.05	10.98	1.61	7.87	0.41	3.40	0.36	4.53
MTBE	8.34	6.86	5.94	2.52	5.87	25.44	9.23	10.52	4.84	3.92	6.63	37.49	13.25	26.54	7.67	2.60	7.19	31.80
α-Pinene	4.33	8.17	1.92	2.89	1.02	31.89	4.01	4.15	2.53	2.59	2.18	15.19	6.53	5.31	4.73	2.31	4.85	17.32
β-Pinene	1.24	1.27	0.88	2.16	0.51	4.93	3.25	5.70	1.42	3.27	1.23	19.53	6.97	6.46	4.55	2.63	4.46	19.88
Styrene	1.22	1.74	0.70	2.51	0.42	6.68	2.22	4.73	0.85	3.07	0.42	16.94	1.65	4.69	0.78	2.57	0.78	3.14
Toluene	15.17	15.24	10.32	2.45	11.10	50.54	26.39	49.34	13.25	2.77	11.27	209.1	16.61	19.85	11.64	2.25	12.58	46.95
TCE	0.28	0.16	0.25	1.45	0.22	0.69	1.01	1.85	0.52	2.69	0.52	7.08	0.16	0.13	0.14	1.54	0.12	0.40
PERC	5.00	15.25	1.82	3.03	1.55	33.95	1.65	2.95	1.03	2.34	0.97	5.75	2.65	18.26	0.46	3.41	0.39	5.76
m,p-Xylene	4.63	3.04	3.55	2.30	4.13	11.59	10.69	31.49	4.80	2.71	3.90	19.90	9.36	11.45	6.05	2.38	5.53	38.98
o-Xylene	1.77	0.89	1.51	1.89	1.72	3.34	3.26	9.12	1.57	2.58	1.48	6.65	3.08	3.40	2.12	2.33	2.01	10.07

n, sample size; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation.

Table 11. Spearman rank correlation coefficients among outdoor VOC measurements in RIOPA.

Outdoor	Benzene	CTC	Chloroform	1,4-DCB	Ethylbenzene	d-Limonene	MC	MTBE	α-Pinene	β-Pinene	Styrene	Toluene	TCE	PERC	m,p-Xylene	o-Xylene
Benzene	1															
CTC	**0.337**	1														
Chloroform	**0.282**	**0.722**	1													
1,4-DCB	0.025	**0.125**	**0.369**	1												
Ethylbenzene	**0.657**	**0.804**	**0.629**	**0.153**	1											
d-Limonene	**0.182**	**0.261**	**0.481**	**0.39**	**0.345**	1										
MC	**0.408**	**0.204**	**0.16**	0.023	**0.393**	**0.082**	1									
MTBE	**0.623**	0.061	**0.145**	-0.017	**0.38**	**0.263**	**0.288**	1								
α-Pinene	**0.174**	**0.431**	**0.533**	**0.116**	**0.419**	**0.468**	**0.101**	**0.251**	1							
β-Pinene	**0.363**	**0.886**	**0.795**	**0.181**	**0.747**	**0.483**	**0.219**	**0.132**	**0.488**	1						
Styrene	**0.43**	**0.95**	**0.753**	**0.135**	**0.866**	**0.313**	**0.276**	**0.179**	**0.483**	**0.891**	1					
Toluene	**0.493**	**0.137**	**0.214**	0.054	**0.405**	**0.41**	**0.378**	**0.547**	**0.285**	**0.206**	**0.216**	1				
TCE	**0.312**	**0.967**	**0.708**	**0.154**	**0.796**	**0.258**	**0.213**	0.047	**0.415**	**0.861**	**0.938**	**0.152**	1			
PERC	**0.516**	**0.809**	**0.613**	**0.115**	**0.837**	**0.348**	**0.464**	**0.263**	**0.421**	**0.758**	**0.843**	**0.36**	**0.804**	1		
m,p-Xylene	**0.757**	**0.504**	**0.442**	**0.079**	**0.815**	**0.354**	**0.548**	**0.63**	**0.345**	**0.521**	**0.625**	**0.599**	**0.51**	**0.733**	1	
o-Xylene	**0.324**	**0.399**	**0.356**	**0.079**	**0.51**	**0.295**	**0.219**	**0.232**	**0.227**	**0.374**	**0.441**	**0.296**	**0.453**	**0.432**	**0.566**	1

Bold type indicates statistically significant ($p<0.05$).

Table 12. Spearman rank correlation coefficients among indoor VOC measurements in RIOPA.

Indoor	Benzene	CTC	Chloroform	1,4-DCB	Ethylbenzene	d-Limonene	MC	MTBE	α-Pinene	β-Pinene	Styrene	Toluene	TCE	PERC	m,p-Xylene	o-Xylene
Benzene	1															
CTC	**0.483**	1														
Chloroform	**0.201**	**0.201**	1													
1,4-DCB	**0.258**	**0.417**	**0.073**	1												
Ethylbenzene	**0.229**	**0.181**	0.063	**0.127**	1											
d-Limonene	0.065	**0.072**	**0.11**	0.053	0.002	1										
MC	-0.007	0.062	0.004	-0.018	0.023	0.019	1									
MTBE	**0.546**	**0.603**	**0.204**	**0.18**	**0.279**	0.042	-0.008	1								
α-Pinene	**0.233**	**0.409**	**0.22**	**0.218**	**0.095**	**0.258**	**0.081**	**0.284**	1							
β-Pinene	**0.282**	**0.262**	**0.239**	**0.097**	0.044	**0.172**	**0.072**	**0.256**	**0.577**	1						
Styrene	**0.092**	0.009	0.04	**0.139**	**0.218**	0.01	**0.435**	0.043	0.07	**0.081**	1					
Toluene	**0.492**	**0.723**	**0.218**	**0.3**	**0.357**	**0.081**	**0.238**	**0.633**	**0.407**	**0.309**	**0.217**	1				
TCE	-0.004	0.036	0.047	0.024	0	0.048	0.022	0.004	0.051	0.014	-0.01	0.044	1			
PERC	**0.157**	**0.29**	**0.096**	**0.084**	**0.135**	-0.015	0.012	**0.229**	**0.107**	0.031	0.038	**0.306**	-0.008	1		
m,p-Xylene	**0.22**	**0.175**	0.043	**0.125**	**0.966**	0.006	0.03	**0.272**	**0.091**	0.046	**0.227**	**0.369**	0.003	**0.092**	1	
o-Xylene	**0.263**	**0.232**	0.06	**0.151**	**0.955**	0.009	0.024	**0.329**	**0.116**	0.065	**0.228**	**0.408**	0.005	**0.116**	**0.98**	1

Bold type indicates statistically significant ($p<0.05$).

Table 13. Spearman rank correlation coefficients among personal adult VOC measurements in RIOPA.

Adult	Benzene	CTC	Chloroform	1,4-DCB	Ethylbenzene	d-Limonene	MC	MTBE	α-Pinene	β-Pinene	Styrene	Toluene	TCE	PERC	m,p-Xylene	o-Xylene
Benzene	**1**															
CTC	**0.554**	1														
Chloroform	**0.667**	0.747	**1**													
1,4-DCB	**0.077**	0.002	**0.012**	**1**												
Ethylbenzene	**0.424**	0.309	**0.248**	**0.05**	**1**											
d-Limonene	**0.62**	0.666	**0.912**	**0.024**	**0.22**	1										
MC	**0.007**	0.009	**0.015**	**-0.022**	**-0.004**	0.019	1									
MTBE	**0.432**	0.116	**0.119**	**0.019**	**0.52**	0.114	-0.01	1								
α-Pinene	**0.122**	0.158	**0.138**	**0.07**	**0.056**	0.202	0.03	0.017	1							
β-Pinene	**0.181**	0.071	**0.018**	**0.083**	**0.024**	0.138	0.002	0.037	0.635	1						
Styrene	**0.092**	0.12	**-0.005**	**0.153**	**0.228**	-0.008	0.134	0.044	0.042	0.058	**1**					
Toluene	**0.607**	0.535	**0.717**	**0.079**	**0.544**	0.657	0.075	0.366	0.112	0.026	**0.108**	1				
TCE	**0.034**	0.139	**0.067**	**-0.016**	**0.021**	0.061	0.004	-0.002	0.049	0.017	**0.011**	0.039	1			
PERC	**0.015**	0.033	**0.035**	**-0.011**	**-0.007**	0.026	-0.004	-0.003	0.024	-0.007	**-0.008**	0.013	0.797	1		
m,p-Xylene	**0.408**	0.269	**0.267**	**0.048**	**0.961**	0.239	0.004	0.446	0.048	0.012	**0.234**	0.567	0.013	0	**1**	
o-Xylene	**0.465**	0.311	**0.281**	**0.037**	**0.952**	0.252	0	0.649	0.063	0.034	**0.212**	0.592	0.025	0	**0.944**	**1**

Bold type indicates statistically significant ($p<0.05$).

Table 14. Spearman rank correlation coefficients among personal child VOC measurements in RIOPA.

Child	Benzene	CTC	Chloroform	1,4-DCB	Ethylbenzene	d-Limonene	MC	MTBE	α-Pinene	β-Pinene	Styrene	Toluene	TCE	PERC	m,p-Xylene	o-Xylene
Benzene	**1**															
CTC	**0.031**	1														
Chloroform	**0.035**	0.228	**1**													
1,4-DCB	**0.225**	0.074	**0.134**	**1**												
Ethylbenzene	**0.073**	0.089	**0.016**	**0.071**	**1**											
d-Limonene	**0.043**	0.051	**0.104**	**0.16**	**-0.018**	1										
MC	**-0.039**	-0.029	**-0.043**	**-0.06**	**-0.027**	-0.021	1									
MTBE	**0.205**	0.158	**0.069**	**-0.013**	**0.195**	-0.022	-0.021	1								
α-Pinene	**0.016**	0.139	**0.15**	**0.134**	**-0.004**	0.289	0.065	0.124	1							
β-Pinene	**0.156**	0.109	**0.104**	**0.249**	**-0.044**	0.344	0.035	-0.022	0.5	1						
Styrene	**0.046**	-0.061	**-0.012**	**0.244**	**0.174**	0.022	0.426	-0.003	0.199	0.163	**1**					
Toluene	**0.041**	0.137	**0.081**	**0.105**	**0.379**	0.138	0.165	0.192	0.069	-0.013	**0.202**	1				
TCE	**-0.066**	0.075	**0.07**	**-0.035**	**0.034**	-0.066	0.016	0.054	0.064	-0.081	**-0.013**	0.105	1			
PERC	**-0.026**	0.035	**-0.028**	**-0.042**	**-0.02**	-0.046	0.026	-0.011	-0.015	0.037	**0.003**	0.03	-0.002	1		
m,p-Xylene	**0.071**	0.119	**0.004**	**0.132**	**0.826**	-0.019	-0.011	0.226	0.047	-0.026	**0.262**	0.497	0.054	-0.01	**1**	
o-Xylene	**0.086**	0.139	**0.01**	**0.154**	**0.761**	-0.008	0.031	0.276	0.099	0	**0.299**	0.527	0.057	-0.006	**0.972**	**1**

Bold type indicates statistically significant ($p<0.05$)

Table 15. Statistics of VOC concentrations (μg L^{-1}) in blood measured for NHANES III and continuous NHANES.

VOCs	NHANES III: 1988-1994							Continuous NHANES: 1999-2004						
	n	DF	Mean	SE	50th	90th	95th	n	DF	Mean	SE	50th	90th	95th
Aromatics														
Benzene	796	66	0.132	0.008	0.062	0.323	0.476	2482	62	0.091	0.006	0.032	0.190	0.320
Toluene	575	56	0.596	0.008	0.281	1.081	1.478	2587	95	0.278	0.014	0.120	0.578	0.880
Ethylbenzene	606	56	0.125	0.004	0.061	0.183	0.245	2439	68	0.049	0.002	0.031	0.089	0.133
m,p-Xylene	1018	62	0.246	0.033	0.117	0.414	0.607	2602	97	0.206	0.012	0.140	0.374	0.512
o-Xylene	628	59	0.153	0.004	0.101	0.198	0.267	2654	41	0.054	0.002	0.035	0.087	0.116
BTEX	1018	NA	0.845	0.101	0.463	1.642	2.380	2703	NA	0.645	0.030	0.363	1.293	1.842
Styrene	624	54	0.094	0.001	0.041	0.129	0.177	2476	52	0.068	0.012	0.021	0.110	0.158
THMs														
Chloroform	876	47	0.042	0.002	0.023	0.072	0.118	2216	95	0.027	0.003	0.014	0.053	0.079
BDCM	937	13	0.008	0.001	0.006	0.011	0.019	2461	86	0.003	0.000	0.002	0.007	0.011
DBCM	919	11	0.010	0.000	0.009	0.015	0.022	2464	64	0.002	0.000	0.001	0.005	0.008
Bromoform	579	4.5	0.021	0.000	0.019	0.019	0.034	2413	60	0.003	0.001	0.001	0.005	0.010
∑THM	1016	NA	0.065	0.003	0.049	0.107	0.147	2513	NA	0.032	0.002	0.018	0.066	0.100
Others														
1,4-DCB	915	86	1.112	0.122	0.322	4.658	11.03	2409	57	0.872	0.102	0.140	1.900	5.300
PERC	566	41	0.219	0.005	0.061	0.347	0.617	2577	29	0.081	0.007	0.034	0.090	0.180
MTBE	NA	NA	NA	NA	NA	NA	NA	2263	85	0.041	0.005	0.013	0.110	0.159

Sample size n includes measurements below MDL, which were replaced by 1/2 MDLs.
Statistical analyses only accounted for detectable measurements and measurements below MDLs, which were replaced by 1/2 MDLs.
DF, detection frequency (%); SE, standard error; NA, not available.

Table 16. Spearman rank correlation coefficients for blood BTEX and THM compounds in NHANES III (top) and continuous NHANES (bottom).

1988-1994* (n = 1338)	Benzene	Toluene	Ethylbenzene	m,p-Xylene	o-Xylene	BTEX	Chloroform	DBCM	BDCM	Bromoform	∑THM
Benzene	1.0										
Toluene	**0.42**	1.00									
Ethylbenzene	**0.23**	**0.59**	1.00								
m,p-Xylene	**0.14**	**0.46**	**0.62**	1.00							
o-Xylene	0.08	**0.38**	**0.81**	**0.49**	1.00						
BTEX	**0.42**	**0.88**	**0.79**	**0.77**	**0.63**	1.00					
Chloroform	**0.09**	-0.01	**0.20**	**0.44**	**0.28**	**0.25**	1.00				
DBCM	-0.01	-0.03	-0.04	-0.05	0.02	-0.03	**0.09**	1.00			
BDCM	0.04	-0.06	-0.05	-0.04	-0.01	-0.06	**0.27**	**0.37**	1.00		
Bromoform	-0.03	-0.10	-0.06	-0.06	-0.04	-0.08	-0.02	**0.14**	**0.36**	1.00	
∑THM	-0.02	-0.01	**0.19**	**0.44**	**0.28**	**0.25**	**0.99**	0.05	0.04	0.04	1.00
1999-2004 (n = 3789)											
Benzene	1.00										
Toluene	**0.76**	1.00									
Ethylbenzene	**0.68**	**0.74**	1.00								
m,p-Xylene	**0.38**	**0.49**	**0.70**	1.00							
o-Xylene	**0.62**	**0.73**	**0.89**	**0.71**	1.00						
BTEX	**0.76**	**0.92**	**0.87**	**0.62**	**0.89**	1.00					
Chloroform	**0.11**	**0.11**	**0.04**	0.04	**0.11**	**0.11**	1.00				
DBCM	0.04	0.03	**0.06**	**0.06**	**0.05**	0.03	**0.11**	1.00			
BDCM	-0.01	0.02	0.00	0.01	0.02	0.00	**0.48**	**0.70**	1.00		
Bromoform	**-0.05**	**-0.10**	0.03	**0.10**	-0.01	**-0.06**	**-0.08**	**0.46**	**0.19**	1.00	
∑THM	0.01	0.02	-0.01	0.01	**0.05**	0.02	**0.90**	**0.36**	**0.59**	**0.22**	1.00

*, excludes 1988-1991 data for toluene, ethylbenzene, m,p-xylene, o-xylene, BTEX, styrene, bromoform, ∑THM and PERC.
Bold type means statistically significant ($p<0.05$).

Table 17. Spearman rank correlations between blood and personal airborne VOCs in NHANES 1999/2000.

Air \ Blood	Benzene	Toluene	Ethylbenzene	m,p-Xylene	o- Xylene	BTEX	Chloroform	1,4-DCB	PERC
Benzene	**0.24**	**0.25**	**0.26**	**0.29**	**0.25**	**0.24**	**-0.17**	-0.06	-0.04
Toluene	**0.15**	**0.26**	**0.23**	**0.24**	**0.24**	**0.21**	**-0.15**	-0.01	0.02
Ethylbenzene	**0.15**	**0.23**	**0.35**	**0.35**	**0.33**	**0.27**	-0.05	-0.04	0.04
m,p-Xylene	**0.16**	**0.25**	**0.36**	**0.38**	**0.35**	**0.28**	-0.04	0.01	0.11
o-Xylene	**0.17**	**0.25**	**0.36**	**0.38**	**0.36**	**0.28**	-0.05	0.02	**0.16**
BTEX	**0.20**	**0.31**	**0.34**	**0.37**	**0.34**	**0.31**	-0.08	0.01	0.04
Chloroform	-0.11	-0.08	-0.11	-0.06	-0.05	**-0.13**	**0.38**	**0.18**	**0.21**
1,4-DCB	-0.08	-0.01	-0.03	0.04	0.01	-0.03	**0.16**	**0.65**	**0.18**
PERC	**-0.27**	**-0.22**	**-0.13**	**-0.13**	-0.07	**-0.22**	**0.22**	**0.17**	**0.62**

Shaded values show correlations for same compounds.
Bold type means statistically significant ($p<0.05$).

Table 18. Identification of best-fit distributions (first rank) for VOCs in RIOPA by sample type.

VOCs	Best-fit distribution								
	Untransformed					Log-transformed			
	Outdoor	Indoor	Adult	Adult_NH	Child	Outdoor	Indoor	Adult	Child
Benzene	Gamma	ExtValue	Pearson5	Lognormal	Pearson5	Normal	Logistic	Logistic	Logistic
Toluene	Logistic	ExtValue	Pearson5	Lognormal	Pearson5	Logistic	Normal	Logistic	Logistic
Ethylbenzene	Gamma	Pearson5	Pearson5	Lognormal	LogLogistic	Weibull	Logistic	Logistic	Logistic
m,p-Xylene	Lognormal	Pearson5	Pearson5	Lognormal	LogLogistic	Logistic	Logistic	Logistic	LogLogistic
o-Xylene	Lognormal	LogLogistic	Pearson5	Lognormal	LogLogistic	Normal	Logistic	Logistic	Logistic
MTBE	Pearson5	Pearson5	Pearson5	Weibull	LogLogistic	Logistic	Logistic	Logistic	Logistic
Styrene	Pearson5	Pearson5	Pearson5	NA	Pearson5	Normal	LogLogistic	Pearson5	LogLogistic
1,4-DCB	Pearson5	Student	Student	Pareto	Logistic	ExtValue	InvGauss	InvGauss	Weibull
MC	LogLogistic	Pearson5	Pearson5	NA	Student	Normal	Logistic	Student	Normal
TCE	Student	Student	Student	Pareto	Student	Logistic	ExtValue	ExtValue	Logistic
PERC	Pearson5	Exponential	Lognormal	Lognormal	InvGauss	Normal	Logistic	Logistic	LogLogistic
Chloroform	Student	Lognormal	Lognormal	Lognormal	Pearson5	ExtValue	Normal	Normal	Logistic
CTC	LogLogistic	LogLogistic	LogLogistic	NA	LogLogistic	Logistic	Logistic	Logistic	Logistic
d-Limonene	Student	Pearson5	Pearson5	NA	Pearson5	ExtValue	Logistic	Logistic	Logistic
α-Pinene	LogLogistic	Lognormal	Lognormal	NA	LogLogistic	Normal	Weibull	Logistic	BetaGeneral
β-Pinene	ChiSq	ExtValue	ExtValue	NA	ExtValue	Normal	Logistic	Logistic	Normal

NA, not available; adult_NH, personal airborne exposures in the 1999/2000 NHANES database.

Table 19. Predicted excess cancer risks for adult participants in RIOPA (n = 239).

VOCs	Unit risk $(\mu g\ m^{-3})^{-1}$	Predicted excess cancer cases per million population									
		Mean	SD	Min	25th	50th	75th	90th	95th	98th	Max
Benzene	7.8×10^{-6}	28.4	25.9	4.3#	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#	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#	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#	0.8#	1.5	2.6	5.8	12.9	23.9	59.9
1,4-DCB	1.1×10^{-5}	626.5	2223	2.4#	10.0#	24.5	126.0	908.9	3620.7	9518.1	19167
TCE	2.0×10^{-6}	1.4	4.1	0.2#	0.2#	0.4#	0.93	2.2	4.6	16.1	40.9
PERC	5.9×10^{-6}	12.9	25.9	0.7#	2.5#	5.9	11.8	24.1	47.1	97.5	242.3
Chloroform	2.3×10^{-5}	47.0	62.2	3.2#	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#	8.2	9.3	10.7	12.9	15.0	17.1	27.8
Hematopoietic mixture	NA	680.2	2240	12.78	44.89	76.4	180.22	965.4	3651.5	9695.8	19196
Liver and kidney toxicant mixture	NA	714.8	2247	20.80	61.25	111.1	265.03	1102.2	3683.6	9723.1	19223
Total VOCs	NA	745.8	2254	34.1	83.9	141.1	293.3	1125.0	3710.1	9780.5	19250

NA, not available; SD, standard deviation.
#, concentrations were based on MDLs.
Hematopoietic mixture includes benzene, MTBE, 1,4-DCB, TCE and PERC; liver and kidney toxicant mixture includes ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform and CTC.

Table 20. Goodness of fit measures (R^2) for the maximum Gumbel distribution fits for 90th and 95th percentile groups in RIOPA by sample type.

VOCs	Outdoor		Indoor		Adult		Adult_NH		Child
	90th%, n=56	95th%, n=28	90th%, n=56	95th%, n=28	90th%, n=54	95th%, n=27	90th%, n=67	95th%, n=33	90th%, n=21
Benzene	0.795	**0.928**	0.788	**0.873**	0.701	0.788	0.79	0.85	0.772
Toluene	0.834	**0.894**	0.706	**0.884**	0.668	0.841	0.61	**0.87**	0.805
Ethylbenzene	0.494	0.639	0.745	**0.916**	0.785	**0.953**	0.38	0.59	0.774
m,p-Xylene	0.703	0.850	0.755	**0.908**	0.776	**0.929**	0.85	**0.95**	0.661
o-Xylene	0.407	0.619	0.742	**0.884**	0.753	**0.908**	0.78	**0.91**	0.682
MTBE	0.790	**0.922**	0.769	**0.915**	0.546	0.718	0.65	0.70	0.651
Styrene	0.358	0.510	0.791	**0.941**	0.808	**0.935**	NA	NA	**0.911**
1,4-DCB	0.430	0.647	**0.884**	**0.965**	**0.912**	**0.950**	0.70	0.79	**0.991**
MC	0.570	0.819	0.586	0.760	0.554	0.758	NA	NA	0.546
TCE	0.284	0.442	0.477	0.715	0.539	0.785	0.62	**0.88**	0.702
PERC	0.512	0.681	0.683	0.793	0.231	0.394	0.45	0.70	0.560
Chloroform	0.524	0.755	0.785	**0.883**	0.227	0.386	**0.89**	**0.94**	0.839
CTC	0.227	0.381	0.407	0.613	0.344	0.546	NA	NA	0.808
d-Limonene	0.837	**0.958**	0.508	0.670	0.407	0.607	NA	NA	0.587
α-Pinene	0.545	**0.867**	**0.870**	**0.977**	0.647	0.802	NA	NA	**0.948**
β-Pinene	0.396	0.686	**0.851**	**0.962**	**0.874**	**0.972**	NA	NA	**0.964**

n, sample size; NA, not available; adult_NH, personal airborne exposures in the 1999/2000 NHANES database.
$R^2 < 0.6$ shows in red, and > 0.85 shows in blue (bold type).

Table 21. GEV parameters and goodness-of-fit for average VOC exposures in RIOPA.

VOCs	Top 10% (n = 24)				Top 5% (n = 12)			
	Shape	Location	Scale	p-value	Shape	Location	Scale	p-value
Benzene	0.4	9.1	2.4	0.876	-0.2	13.6	3.6	0.684
Toluene	1.6	35.8	7.3	0.672	0.6	63.6	19.2	0.829
Ethylbenzene	1.2	6.3	1.7	0.951	0.8	10.6	3.9	0.943
m,p-Xylene	0.8	19.9	6.6	0.963	1.2	28.7	6.9	0.905
o-Xylene	0.9	6.8	2.1	0.900	1.8	10.0	1.3	0.915
MTBE	0.6	36.3	12.5	0.988	0.9	53.0	11.4	0.958
Styrene	1.3	3.9	1.6	0.676	0.9	8.4	2.8	0.895
1,4-DCB	0.5	258.0	188.0	0.991	0.5	516.0	234.9	0.953
TCE	1.1	1.7	0.8	0.987	1.7	2.8	1.0	0.909
PERC	1.0	5.9	2.6	0.882	0.7	11.4	4.2	0.988
Chloroform	0.7	5.5	1.6	0.954	1.1	7.6	1.7	0.943
CTC	0.7	0.9	0.1	0.854	0.7	1.1	0.1	0.991
d-Limonene	0.6	85.8	20.0	0.725	0.4	124.8	19.7	0.890
α-Pinene	1.1	18.0	4.0	0.959	1.7	23.4	6.0	0.797
β-Pinene	0.9	18.2	6.5	0.897	0.1	35.2	13.8	0.905

p-values shown for Anderson-Darling tests.
p-value > 0.05 indicating that observations fit to generalized extreme value distributions.

Table 22. Comparison of adult VOC distributions between observed data and GEV, Gumbel and lognormal simulation in RIOPA using Kolmogorov-Smirnov tests.

VOCs	GEV simulation				Gumbel simulation				Lognormal simulation			
	Top 10%		Top 5%		Top 10%		Top 5%		Top 10%		Top 5%	
	Statistics	p-value	Statistics	p-value	Statistics	p-value	Statistics	p-value	Statistics	p-value	Statistics	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

Sample size of observed data is 239; sample size of simulated data is 10,000.
p-value < 0.05 shown in bold type; p-value > 0.05 indicating that there is no significance difference between two distributions.

Table 23. Evaluation of simulated VOC concentrations above the 90th and 95th percentiles in RIOPA fit to GEV, Gumbel and lognormal distributions.

VOCs	% exceeding	Predicted cancer risks									
		Above the 90th percentile of exposure					Above the 95th percentile of exposure				
		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	Observed measurements	100	100	29	0	0	100	100	58	0	0
	GEV simulation	100	100	26	0	0	100	100	**71**	0	0
	Gumbel simulation	100	100	31	0	0	100	100	67	0	0
	Lognormal simulation	100	100	**18**	0	0	100	100	**35**	0	0
Ethylbenzene	Observed measurements	100	100	0	0	0	100	100	0	0	0
	GEV simulation	100	100	7	1	0	100	100	8	0	0
	Gumbel simulation	100	91	0	0	0	100	98	1	0	0
	Lognormal simulation	100	100	0	0	0	100	100	0	0	0
MTBE	Observed measurements	100	63	0	0	0	100	100	0	0	0
	GEV simulation	100	57	1	0	0	100	100	3	0	0
	Gumbel simulation	98	**74**	0	0	0	99	**87**	0	0	0
	Lognormal simulation	100	**53**	0	0	0	100	100	0	0	0
Styrene	Observed measurements	100	54	0	0	0	100	100	0	0	0
	GEV simulation	100	46	6	1	0	100	100	5	0	0
	Gumbel simulation	96	**69**	0	0	0	100	93	0	0	0
	Lognormal simulation	100	**28**	0	0	0	100	**55**	0	0	0
1,4-DCB	Observed measurements	100	100	100	88	13	100	100	100	100	25
	GEV simulation	100	100	100	96	13	100	100	100	100	27
	Gumbel simulation	96	96	95	89	7	100	100	100	99	24
	Lognormal simulation	100	100	100	**65**	5	100	100	100	100	**10**
TCE	Observed measurements	100	21	0	0	0	100	42	0	0	0
	GEV simulation	100	18	2	0	0	100	33	7	2	0
	Gumbel simulation	**77**	**61**	1	0	0	**83**	**74**	9	0	0
	Lognormal simulation	100	**2**	0	0	0	100	**3**	0	0	0
PERC	Observed measurements	100	100	17	0	0	100	100	33	0	0
	GEV simulation	100	100	18	2	0	100	100	32	1	0
	Gumbel simulation	99	96	16	0	0	100	100	**44**	0	0
	Lognormal simulation	100	100	8	0	0	100	100	**16**	0	0
Chloroform	Observed measurements	100	100	88	0	0	100	100	100	0	0
	GEV simulation	100	100	93	2	0	100	100	100	6	1
	Gumbel simulation	100	100	86	0	0	100	100	98	0	0
	Lognormal simulation	100	100	93	0	0	100	100	100	0	0
CTC	Observed measurements	100	100	0	0	0	100	100	0	0	0
	GEV simulation	100	100	0	0	0	100	100	1	0	0
	Gumbel simulation	96	**81**	0	0	0	**89**	**78**	4	0	0
	Lognormal simulation	100	100	0	0	0	100	100	0	0	0
Hematopoietic mixture	Observed measurements	100	100	100	96	17	100	100	100	100	33
	GEV simulation	100	100	100	97	14	100	100	100	100	27
	Gumbel simulation	97	97	96	90	10	100	100	100	99	30
	Lognormal simulation	100	100	100	**79**	**2**	100	100	100	100	**4**
Liver and kidney toxicant mixture	Observed measurements	100	100	100	100	17	100	100	100	100	33
	GEV simulation	100	100	100	97	14	100	100	100	100	26
	Gumbel simulation	97	97	97	91	10	100	100	100	99	31
	Lognormal simulation	100	100	100	**88**	**1**	100	100	100	100	**3**
Total VOCs	Observed measurements	100	100	100	100	17	100	100	100	100	33
	GEV simulation	100	100	100	98	13	100	100	100	100	27
	Gumbel simulation	97	97	96	92	11	100	100	100	100	32
	Lognormal simulation	100	100	100	97	**1**	100	100	100	100	**1**

Table 24. GEV parameters and goodness-of-fit for the originally weighted personal VOC exposures in NHANES 1999/2000.

VOCs μg m^{-3}	Top 10% (n = 1442 - 1467)					Top 5% (n = 726 - 775)				
	Shape	Location	Scale	p-value for A-D test	p-value for K-S test	Shape	Location	Scale	p-value for A-D test	p-value for K-S test
Benzene	0.42	17	4.3	< 0.05	< 0.05	0.41	23.4	4.3	< 0.05	**0.24**
Toluene	0.82	89.4	35.3	< 0.05	< 0.05	1.29	125.8	51.8	< 0.05	< 0.05
Ethylbenzene	0.94	21.1	9	< 0.05	< 0.05	1.07	35.6	15.1	< 0.05	< 0.05
m,p-Xylene	0.74	62.6	30.1	< 0.05	< 0.05	0.54	117.5	46.4	< 0.05	< 0.05
o-Xylene	0.56	23.2	9.7	< 0.05	< 0.05	0.68	36	11.9	< 0.05	< 0.05
MTBE	0.81	16.7	7.3	< 0.05	< 0.05	0.99	27.6	9.6	< 0.05	< 0.05
1,4-DCB	0.87	88.3	69.8	< 0.05	< 0.05	0.56	234.1	96.2	< 0.05	< 0.05
TCE	1.35	4.4	5.1	< 0.05	< 0.05	1.02	17.1	13	< 0.05	< 0.05
PERC	1.13	12	7.7	< 0.05	< 0.05	0.94	28.2	12.4	< 0.05	< 0.05
Chloroform	0.35	9.7	3.8	< 0.05	< 0.05	0.53	14.5	3	< 0.05	< 0.05

A-D tests were the goodness-of-fit tests for GEV distribution fitting.
K-S tests were used to compare the observations (the whole weighted sample without ties, n = 14,320 to 14,524) with simulated data based on the GEV parameters.
p-value > 0.05 indicating that observations fit to GEV distributions or indicating that the observational measurements were not different from GEV simulations.

Table 25. GEV parameters and goodness-of-fit for the weighted personal VOC exposures that used bootstrap methods and repeated sampling in NHANES 1999/2000.

VOCs μg m^{-3}	Top 10% (n = 64)					Top 5% (n = 32)				
	Shape	Location	Scale	p-value for A-D test	p-value for K-S test	Shape	Location	Scale	p-value for A-D test	p-value for K-S test
Benzene	0.48	16.87	4.18	< 0.05	< 0.05	0.53	23.0	4.0	< 0.05	< 0.05
Toluene	1.07	91.66	42.12	< 0.05	< 0.05	1.80	151.2	111.6	< 0.05	< 0.05
Ethylbenzene	1.02	20.65	8.83	< 0.05	< 0.05	1.26	36.0	17.6	< 0.05	< 0.05
m,p-Xylene	0.88	62.11	27.51	< 0.05	< 0.05	0.54	120.4	45.6	< 0.05	< 0.05
o-Xylene	0.69	22.86	8.85	< 0.05	< 0.05	0.77	36.4	10.9	< 0.05	< 0.05
MTBE	0.92	16.22	6.76	< 0.05	< 0.05	1.06	27.3	9.6	< 0.05	< 0.05
1,4-DCB	0.99	91.37	73.84	< 0.05	< 0.05	0.73	233.7	106.7	**> 0.05**	< 0.05
TCE	1.54	4.49	5.28	< 0.05	< 0.05	1.22	16.9	13.5	**> 0.05**	< 0.05
PERC	1.08	12.37	8.01	< 0.05	< 0.05	1.05	28.1	13.2	< 0.05	< 0.05
Chloroform	0.48	9.42	3.43	< 0.05	< 0.05	0.56	14.6	3.0	< 0.05	< 0.05

A-D tests were the goodness-of-fit tests for GEV distribution fitting using the repeated datasets (n = 635 to 648, 300 times) randomly selected from the weighted samples; values of parameters were averages of 300 results.
K-S tests were used to compare the observations (the whole weighted sample without ties, n = 14,320 to 14,524) with simulated data based on the GEV parameters, which were estimated from the 300 random samples.
p-values were estimated from empirical distributions of statistics, i.e., comparing the observational statistics with the statistics of random samples (repeatedly sampling 300 times); p-value > 0.05 indicating that observations fit to GEV distributions or indicating that the observational measurements were not different from GEV simulations.

Table 26. GEV parameters and goodness-of-fit for the unweighted personal VOC exposures in NHANES 1999/2000.

VOCs μg m^{-3}	Top 10% (n = 64)					Top 5% (n = 32)				
	Shape	Location	Scale	p-value for A-D test	p-value for K-S test	Shape	Location	Scale	p-value for A-D test	p-value for K-S test
Benzene	0.69	15.5	3.7	**0.82**	**0.70**	0.64	21.8	4.4	**0.99**	**0.90**
Toluene	1.1	78.5	33.4	**0.92**	**0.82**	1.76	119.5	43.5	**0.75**	**0.56**
Ethylbenzene	0.93	17.9	8.6	**0.90**	**0.94**	0.87	32.9	14.2	**1.00**	**1.00**
m,p-Xylene	1.18	47.7	20.2	**0.45**	**0.53**	0.57	101.7	47.1	**0.81**	**0.71**
o-Xylene	1.08	17.3	7.4	**0.42**	**0.41**	0.84	32.5	12.4	**0.76**	**0.35**
MTBE	0.86	20.3	8.9	**0.90**	**0.98**	0.94	34.7	11.9	**0.91**	**0.94**
1,4-DCB	0.69	199.4	111.6	**1.00**	**1.00**	1	350.3	122.1	**0.85**	**0.85**
TCE	1.65	5.2	7.1	**0.63**	**0.81**	1.11	22.3	20.6	**0.89**	**0.92**
PERC	1.29	11	6.4	**0.49**	**0.43**	1.16	25.2	10	**0.98**	**0.97**
Chloroform	0.67	8.9	3	**0.63**	**0.31**	0.73	13.7	3	**0.96**	**0.96**

A-D tests were the goodness-of-fit tests for GEV distribution fitting.
K-S tests were used to compare the observations (the whole unweighted sample) with simulated data based on the GEV parameters.
p-value > 0.05 indicating that observations fit to GEV distributions or indicating that the observational measurements were not different from GEV simulations.

Table 27. Goodness of fit statistics of each density estimation method for chloroform, 1,4-DCB and styrene sample data from the RIOPA study.

VOCs	(estimated) Proportion below MDL				MSE			MAE		
	Observed	Normal	MN	DPMN	Normal	MN	DPMN	Normal	MN	DPMN
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

MSE, mean squared error; MAE, mean absolute error; MN, mixture of normals; DPMN, Dirichlet process mixture of normals.
MSE and MAE are multiplied by a scalar of 1,000 to reflect the significant figure.

Table 28. Fitted weight, location and dispersion parameters under the finite mixture of normals for chloroform, 1,4-DCB and styrene sample data from the RIOPA study.

	Chloroform			1,4-DCB			Styrene		
	Weight	Mean	SD	Weight	Mean	SD	Weight	Mean	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

SD, standard deviation.
The smallest AIC shown in bold type.

Table 29. Posterior distribution of the number of clusters K based on various prior settings of α as a sensitivity analysis.

Prior	Posterior distribution of K								
	Chloroform			1,4-DCB			Styrene		
	mean	median	SD	mean	median	SD	mean	median	SD
Setting 1	2.8	2	1.4	32.8	34	20.2	10.9	5	10.8
Setting 2	3.9	3	2.4	5.6	5	2.5	4.6	4	2.8
Setting 3	4.1	4	2.2	7.1	7	3.4	7.9	7	4.4
Setting 4	10.5	9	6.0	15.3	14	6.5	13.1	12	6.0

SD, standard deviation.
Setting 1: $\alpha \sim \text{Gamma}(0.3, 0.4)$; Setting 2: $\alpha \sim \text{Gamma}(1.2, 2.5)$;
Setting 3: $\alpha \sim \text{Gamma}(2, 1.5)$; Setting 4: $\alpha \sim \text{Gamma}(5, 2)$.

Table 30. Summary of goodness of fit statistics of each density estimation method in the simulation study.

	Proportion below MDL	MSE			MAE		
		Normal	MN	DPMN	Normal	MN	DPMN
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

MSE, mean squared error; MAE, mean absolute error; MN, mixture of normals; DPMN, Dirichlet process mixture of normals.
MSE and MAE are multiplied by a scalar of 1000 to reflect the significant figure.
Distribution 1: $\text{Normal}(0, 2^2)$; Distribution 2: $\frac{1}{2}\text{Gamma}(3, 1.5) + \frac{1}{2}\text{Uniform}(-3, 8)$.
Prior distribution on α is $\text{Gamma}(1.2, 2.5)$.

Table 31. Linear quantile regressions of log-transformed blood VOC concentrations for the NHANES 1988 to 2004 period.*

VOCs	0.5 Quantile		0.75 Quantile		0.95 Quantile	
	Slope	SE	Slope	SE	Slope	SE
Aromatics						
Benzene	**-0.054**	0.003	**-0.078**	0.009	**-0.043**	0.025
Toluene	**-0.099**	0.009	**-0.144**	0.017	**-0.118**	0.024
Ethylbenzene	**-0.060**	0.005	**-0.066**	0.008	**-0.103**	0.023
m,p-Xylene	**-0.033**	0.006	**-0.057**	0.008	**-0.117**	0.042
o-Xylene	**-0.069**	0.004	**-0.097**	0.007	**-0.122**	0.028
BTEX	**-0.066**	0.006	**-0.080**	0.010	**-0.071**	0.027
Styrene	**-0.036**	0.004	**-0.039**	0.009	-0.061	0.033
THMs						
Chloroform	**-0.065**	0.005	**-0.064**	0.006	**-0.103**	0.025
BDCM	**-0.097**	0.007	**-0.043**	0.003	**-0.034**	0.012
DBCM	**-0.202**	0.014	**-0.149**	0.005	**-0.077**	0.007
Bromoform	**-0.241**	0.001	**-0.201**	0.000	**-0.128**	0.022
∑THM	**-0.115**	0.007	**-0.101**	0.009	**-0.115**	0.030
Others						
1,4-DCB	**-0.063**	0.001	**-0.045**	0.009	-0.032	0.025
PERC	0.001	0.001	**-0.166**	0.006	**-0.177**	0.042

*, excludes 1988-1991 data for toluene, ethylbenzene, m,p-xylene, o-xylene, BTEX, styrene, bromoform, ΣTHM and PERC.
Aromatic VOCs were adjusted for solvent-related occupations and serum cotinine levels; THMs and other VOCs were adjusted for solvent-related occupations only.
SE, standard error.
Bold type means statistically significant ($p < 0.05$); benzene at 0.95 quantile is borderline significant.

Table 32. Relative changes (%) per year in untransformed blood VOC concentrations in NHANES at various quantiles.*

VOCs	1988-1991 vs. 2003/2004			1988-1991 vs. 1999/2000			1999/2000 vs. 2003/2004		
	0.5	0.75	0.95	0.5	0.75	0.95	0.5	0.75	0.95
Aromatics									
Benzene	**-3.8**	**-4.3**	**-3.3**	5.2	-1.2	0.1	**-18.2**	**-14.8**	**-12.4**
Toluene	**-5.6**	**-5.7**	**-4.7**	-1.9	-2.9	-1.3	**-15.3**	**-14.8**	**-12.9**
Ethylbenzene	**-4.2**	**-4.9**	**-4.9**	-4.2	-4.2	-3.2	**-6.5**	**-9.4**	**-11.1**
m,p-Xylene	**-2.5**	**-3.5**	**-4.5**	**-0.8**	**-1.8**	**-2.4**	**-6.3**	**-7.9**	**-10.8**
o-Xylene	**-5.5**	**-5.5**	**-5.6**	-7.9	-6.8	-5.3	**-2.1**	**-6.4**	**-10.9**
BTEX	**-4.4**	**-4.6**	**-4.2**	**-2.1**	**-1.6**	**-0.8**	**-10.8**	**-12.0**	**-11.9**
Styrene	**-3.9**	**-2.7**	-3.1	0.5	0.1	0.6	**-12.3**	**-8.2**	**-10.1**
THMs									
Chloroform	**-3.9**	**-3.6**	**-3.9**	3.5	6.3	2.5	**-17.5**	**-18.1**	**-17.0**
BDCM	**-5.0**	**-5.6**	**-3.4**	-6.5	-3.5	-3.0	**-3.1**	**-2.6**	**-6.8**
DBCM	**-6.3**	**-3.0**	**-4.8**	-8.1	-6.6	-5.8	**-13.5**	**-10.8**	**-5.9**
Bromoform	**-7.9**	**-7.5**	**-7.0**	-11.9	-11.2	-10.6	**2.8**	**-1.0**	1.1
∑THM	**-5.9**	**-5.4**	**-3.8**	**-5.4**	**-2.1**	2.7	**-12.2**	**-14.4**	**-13.9**
Others									
1,4-DCB	**-3.5**	**-3.7**	-3.7	**-2.3**	-2.5	-4.8	**-9.0**	**-9.8**	-0.8
PERC	-3.2	**-6.2**	**-6.4**	**-2.7**	-4.8	-5.1	-5.3	**-14.8**	**-15.4**

*, excludes 1988-1991 data for toluene, ethylbenzene, m,p-xylene, o-xylene, BTEX, styrene, bromoform, ΣTHM and PERC.
Relative changes of VOC levels per year between study period at each percentile.
Bold type means statistically significant trend ($p<0.05$).

Table 33. Total emissions and relative changes per year of VOCs in NATA.

VOCs	Total emissions (T yr^{-1})			Relative change per year (%)	
	1996	1999	2002	1996 vs. 1999	1999 vs. 2002
Aromatics					
Benzene	337,000	350,776	410,219	1.4	5.6
Toluene	NA	996,443	884,066	NA	-3.8
Ethylbenzene	NA	NA	127,742	NA	NA
o,m,p-Xylene	NA	712,084	595,241	NA	-5.5
Styrene	NA	NA	49,795	NA	NA
THMs					
Chloroform	3,310	15,139	6,805	119.1	-18.3
Bromoform	NA	NA	22	NA	NA
Others					
1,4-DCB	NA	12,794	7,244	NA	-14.5
PERC	44,100	46,793	35,613	2.0	-8.0

NA, not available.

Table 34. Ambient concentrations and change per year of aromatics for various concentration quantiles in PAMS.

VOCs	Ambient concentrations (ppb)								Relative change per year (%)		
	1993	1994	1999	2000	2001	2002	2003	2004	1993-1999	1999-2004	1993-2004
Aromatics	Mean										
Benzene	3.43	4.18	2.33	1.96	0.99	0.86	0.74	0.70	-5.3	-14.0	-7.2
Toluene	8.12	9.85	5.99	4.85	2.89	2.52	2.33	2.30	-4.4	-12.3	-6.5
Ethylbenzene	2.07	1.87	1.13	0.96	0.49	0.46	0.38	0.34	-7.5	-14.1	-7.6
m,p-Xylene	5.36	5.37	3.34	2.57	1.33	1.17	0.93	0.80	-6.3	-15.2	-7.7
o-Xylene	2.48	2.18	1.34	1.08	0.56	0.53	0.43	0.38	-7.7	-14.4	-7.7
Styrene	1.77	1.17	0.57	0.58	0.26	0.27	0.31	0.41	-11.3	-5.7	-7.0
Aromatics	0. 5 Quantile										
Benzene	1.80	2.21	1.18	1.00	0.47	0.41	0.23	0.14	-5.7	-17.6	-8.4
Toluene	4.10	4.77	2.82	2.27	0.87	0.67	0.35	0.26	-5.2	-18.2	-8.5
Ethylbenzene	0.90	0.97	0.59	0.48	0.15	0.10	0.05	0.03	-5.7	-19.1	-8.8
m,p-Xylene	2.70	2.65	1.46	1.06	0.31	0.25	0.10	0.05	-7.7	-19.3	-8.9
o-Xylene	1.10	1.10	0.64	0.50	0.17	0.12	0.05	0.03	-6.9	-19.2	-8.9
Styrene	0.60	0.56	0.33	0.30	0.11	0.10	0.10	0.04	-7.6	-17.8	-8.5
Aromatics	0.75 Quantile										
Benzene	3.70	4.40	2.23	2.00	1.26	1.08	0.90	0.75	-6.6	-13.3	-7.2
Toluene	9.29	10.70	6.06	5.09	3.32	2.70	2.40	1.82	-5.8	-14.0	-7.3
Ethylbenzene	1.90	2.00	1.20	1.01	0.60	0.51	0.43	0.36	-6.1	-14.0	-7.4
m,p-Xylene	6.00	5.86	3.36	2.59	1.47	1.17	0.94	0.74	-7.3	-15.6	-8.0
o-Xylene	2.41	2.40	1.40	1.10	0.66	0.58	0.47	0.39	-7.0	-14.4	-7.6
Styrene	1.81	1.27	0.58	0.56	0.33	0.32	0.34	0.32	-11.3	-9.0	-7.5
Aromatics	0.95 Quantile										
Benzene	11.58	14.80	7.31	6.30	3.62	3.20	3.00	2.84	-6.1	-12.2	-6.9
Toluene	31.39	35.50	20.69	17.06	11.90	10.44	10.71	9.30	-5.7	-11.0	-6.4
Ethylbenzene	6.20	6.65	3.70	3.36	2.02	1.87	1.69	1.51	-6.7	-11.8	-6.9
m,p-Xylene	20.50	19.64	11.66	9.81	5.68	4.88	4.30	3.80	-7.2	-13.5	-7.4
o-Xylene	7.90	7.89	4.50	4.15	2.30	2.20	2.00	1.74	-7.2	-12.3	-7.1
Styrene	8.04	3.21	1.66	1.70	0.90	1.02	1.07	1.55	-13.2	-1.3	-7.3

Table 35. Site-weighted average of average concentrations (ppb) of ambient VOCs in AQS datasets.*

Year	Benzene		Toluene		Ethylbenzene		m,p-Xylene		o-Xylene		Styrene		Chloroform		1,4-DCB		PERC	
	n	mean	n	mean	n	mean	n	mean	n	mean	n	mean	n	mean	n	mean	n	mean
1990	590	7.92	530	23.27	NA	NA	NA	NA	NA	NA	NA	NA	614	0.727	NA	NA	NA	NA
1991	754	6.65	582	18.09	NA	NA	NA	NA	NA	NA	NA	NA	779	0.670	NA	NA	773	0.55
1992	1066	5.54	751	13.82	NA	NA	NA	NA	NA	NA	NA	NA	1189	0.399	NA	NA	1184	0.37
1993	1318	6.21	1087	14.53	888	3.208	NA	NA	824	3.23	915	1.47	1581	0.179	NA	NA	1578	0.37
1994	1600	6.37	1367	13.43	1164	3.197	949	7.83	1103	2.73	1229	2.47	1839	0.153	NA	NA	1844	0.39
1995	1981	5.51	1673	11.91	1413	2.201	1399	6.13	1351	2.20	1465	2.05	2050	0.110	919	0.53	2043	0.30
1996	2224	4.14	1908	12.50	1672	1.568	1692	5.61	1642	1.75	1722	4.00	2367	0.051	976	0.55	2357	0.21
1997	2491	4.68	2164	12.02	1923	1.957	1918	5.48	1863	2.35	2007	1.49	2727	0.053	943	0.83	2763	0.42
1998	2991	3.99	2502	8.70	2380	1.701	2381	5.41	2197	1.77	2202	0.93	3004	0.058	560	0.61	2995	0.26
1999	3586	3.99	3171	8.00	3029	1.252	3050	5.16	2965	1.75	3010	0.73	3529	0.054	1069	0.38	3535	0.11
2000	4407	3.61	3976	7.52	3845	1.434	3769	6.17	3622	2.18	3568	1.38	4630	0.055	1850	0.80	4787	0.11
2001	5307	4.00	5005	7.87	4637	1.227	4603	4.38	4446	1.56	4456	0.81	5689	0.048	2704	0.69	5805	0.12
2002	6860	3.11	6509	6.55	6044	1.077	6075	3.46	5973	1.43	6015	0.60	6962	0.059	3988	0.56	7041	0.15
2003	8106	2.98	7674	5.54	7717	0.973	7643	2.62	7581	1.05	7283	2.66	8230	0.063	5009	0.46	8264	0.17
2004	9507	2.59	9054	4.51	9155	0.828	8648	2.10	8864	0.85	8504	0.54	9791	0.069	6254	0.38	9894	0.18
% change[1]	**-4.8**		**-5.8**		**-6.7**		**-7.3**		**-6.7**		**-5.7**		**-19.7, -4.2**		**-3.2**		**-5.2**	
% change[2]	**-4.7**		**-5.7**		**-7.1**		**-6.4**		**-5.8**		**-5.5**		**-21.1, NA**		**-2.8**		**-6.7**	

*, AQS data used 24 h sampling, 24 or more measurements per site-year in EPA Region 1-10.
n, number of observations; NA, not available.
BTEX observations in Site 42 (Edinburg), Region 6 in 1997 were excluded due to extremely high values.
% change[1], relative change per year from the beginning year to 2004. For benzene and toluene, the beginning year is 1990; for ethylbenzene, o-xylene, and styrene, the beginning year is 1993; for m,p-xylene, the beginning year is 1994; for 1,4-DCB, the beginning year is 1995; for PERC, the beginning year is 1991. The relative changes per year for chloroform were calculated from 1990 to 1994 (-19.7%) and 1995 to 2004 (-4.2%).
% change[2], relative change per year were estimated by regression models from the beginning year to 2004. For example, the estimated relative change for benzene is (((benzene$_{2004}$- benzene$_{1990}$)/ benzene$_{1990}$)*100%)/(2004-1990). Since the estimated chloroform levels in 1995-2004 were negative, the relative change did not be calculated.

Table 36. Modeled ambient concentrations and relative changes per year of VOCs for mean and two quantiles in NATA.

VOCs	Modeled ambient concentrations (μg m^{-3})							Relative change per year (%)			
	Mean			0.5		0.95		Mean		0.5	0.95
	1996	1999	2002	1996	1999	1996	1999	1996 vs. 1999	1999 vs. 2002	1996 vs. 1999	1996 vs. 1999
Aromatics											
Benzene	1.39	1.37	1.21	1.21	1.16	2.84	3.12	-0.5	-3.9	-1.4	3.3
Toluene	NA	3.02	2.54	NA	2.21	NA	8.61	NA	-5.2	NA	NA
Ethylbenzene	NA	NA	0.28	NA	NA	NA	NA	NA	NA	NA	NA
o,m,p-Xylene	NA	2.23	1.25	NA	1.60	NA	6.60	NA	-14.7	NA	NA
Styrene	NA	NA	0.05	NA	NA	NA	NA	NA	NA	NA	NA
THMs											
Chloroform	0.09	0.09	0.09	0.08	0.07	0.11	0.21	0.9	-1.7	-5.5	32.4
Bromoform	NA	NA	0.00	NA	NA	NA	NA	NA	NA	NA	NA
Others											
1,4-DCB	NA	0.06	0.06	NA	0.03	NA	0.21	NA	0.1	NA	NA
PERC	0.32	0.31	0.15	0.24	0.19	0.78	1.14	-1.5	-17.5	-7.9	15.4

NA, not available.

Table 37. Sources and apportionments of outdoor, indoor, and personal VOCs (non-averaged measurements) in RIOPA by sample type and seasons, based on PMF results.

Type	Season	Factor	Source Category	Included VOCs	Apportionment	
					%	μg m^{-3}
Outdoor	Warm	1	Gasoline	MTBE	32	8.9
		2	Vehicle exhaust and industrial sources	Aromatics, TCE, chloroform, CTC and β-pinene	32	8.9
		3	Cleaning products and odorants	1,4-DCB and D-limonene	18	5.1
		4	Industrial and biogenic sources	Styrene, 1,4-DCB, TCE, PERC, chloroform, CTC and α-pinene	18	4.9
	Cold	1	Vehicle exhaust	BTEX	34	11.7
		2	Gasoline	MTBE and toluene	27	9.2
		3	Cleaning products, odorants and industrial sources	Styrene, 1,4-DCB, TCE, chloroform, CTC, α-pinene, β-pinene and D-limonene	22	7.6
		4	Industrial and biogenic sources	Styrene, PERC and α-pinene	17	5.9
Indoor	Warm	1	Moth repellents and odorants	1,4-DCB	52	85.3
		2	Cleaning products and odorants	D-limonene, α-pinene and β-pinene	21	35.1
		3	Vehicle exhaust, chlorinated solvents, and cleaning products	Aromatics, TCE, PERC, chloroform, CTC, α-pinene and β-pinene	14	23.8
		4	Gasoline	Benzene and MTBE	13	21
	Cold	1	Moth repellents and odorants	1,4-DCB	39	52.5
		2	Cleaning products and odorants	D-limonene, α-pinene and β-pinene	26	35.3
		3	Vehicle exhaust, chlorinated solvents, and cleaning products	Aromatics, TCE, PERC, chloroform, CTC, α-pinene and β-pinene	21	27.6
		4	Gasoline	MTBE	14	18
Personal	Warm	1	Cleaning products and odorants	D-limonene, α-pinene and β-pinene	42	42.3
		2	Vehicle exhaust	Ethylbenzene, m,p-xylene and o-xylene	22	22.6
		3	Gasoline	Benzene and MTBE	20	19.8
		4	Moth repellents and chlorinated solvents	1,4-DCB, TCE, PERC, chloroform and CTC	15	15.3
	Cold	1	Cleaning products and odorants	D-limonene, α-pinene and β-pinene	44	45.1
		2	Gasoline, chlorinated solvents, and cleaning products	Benzene, toluene, MTBE, styrene, 1,4-DCB, TCE, chloroform and CTC	27	27.2
		3	Vehicle exhaust	Ethylbenzene, m,p-xylene and o-xylene	20	19.9
		4	Dry cleaning solvent	PERC	7.7	7.8

Personal measurements include adult and child exposure data.
Warm season indicates April to September, and cold season indicates October to March.
Apportionment indicates source contributions to the total VOCs by the percentages and concentrations.

Table 38. Sources and apportionments of mixtures of VOCs derived using PMF and the first-visit measurements in RIOPA.

Mixture ID	Suggested Source Categories	VOC Components	Fraction of TVOC	
			%	μg m^{-3}
A1	Gasoline	Benzene and MTBE	20.5	19.9
A2	Vehicle exhaust	Toluene, ethylbenzene, xylenes, and styrene	20.9	20.3
A3	Moth repellents, chlorinated solvents and disinfection by-products	1,4-DCB, TCE, PERC, chloroform, and CTC	16.3	15.9
A4	Cleaning products and odorants	d-Limonene, α-pinene, and β-pinene	42.3	41.1

Table 39. Results of bivariate logistic regression models for VOC mixtures identified by PMF analyses in RIOPA.

Potential factor		Mixtures							
		Benzene and MTBE		Toluene, ethylbenzene, xylenes, and styrene		1,4-DCB, TCE, PERC, chloroform, and CTC		d-Limonene, α-pinene, and β-pinene	
Categorical variables	Group	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
City	CA	**0.56**	**0.35-0.90**	**0.45**	**0.28-0.73**	**0.49**	**0.30-0.79**	**0.18**	**0.11-0.30**
	NJ	**0.39**	**0.24-0.63**	**0.51**	**0.31-0.82**	0.63	0.39-1.03	**0.21**	**0.12-0.34**
	TX	Reference		Reference		Reference		Reference	
Ethnicity	Mexican	**2.03**	**1.19-3.47**	1.57	0.92-2.67	**3.21**	**1.87-5.54**	**3.97**	**2.29-6.87**
	Hispanic	1.07	0.66-1.75	1.35	0.82-2.20	**1.78**	**1.09-2.92**	0.98	0.60-1.61
	Other	0.58	0.30-1.12	**0.47**	**0.24-0.92**	1.66	0.86-3.21	0.86	0.45-1.66
	White	Reference		Reference		Reference		Reference	
Employment	Yes	0.95	0.63-1.42	0.98	0.65-1.47	1.02	0.68-1.52	**0.40**	**0.27-0.61**
	No	Reference		Reference		Reference		Reference	
Attached garage	Yes	**2.27**	**1.45-3.56**	**1.95**	**1.25-3.05**				
	No	Reference		Reference					
Open doors or windows	Yes	0.79	0.52-1.18	**0.40**	**0.26-0.61**	**0.36**	**0.24-0.55**	**0.32**	**0.21-0.49**
	No	Reference	Reference	Reference	Reference				
Self-service pump gas	Yes	**2.10**	**1.25-3.52**	1.62	0.97-2.70				
	No	Reference		Reference					
Other family members take showers	Yes					**2.06**	**1.20-3.56**	**2.45**	**1.42-4.23**
	No						Reference		Reference
Use fresheners	Yes					1.37	0.73-2.57	**2.20**	**1.17-4.14**
	No						Reference		Reference
Continuous variables	Unit								
Log-transformed AERs	hr^{-1}	**0.69**	**0.54-0.89**	**0.45**	**0.35-0.58**	**0.49**	**0.38-0.63**	**0.38**	**0.29-0.49**

OR, odds ratio; CI, confidence interval.
Statistically significant ORs are shown in bold type.

Table 40. Observed and estimated probability of high concentration mixtures in RIOPA.

Mixture ID	VOCs	Copula	Percentile	Probability		
				Observed (n = 299)	Uncorrelated	Copula (n = 1,000)
A1	Benzene and MTBE	Gumbel	50th	0.3545	0.2500	0.3470
			75th	0.1371	0.0625	0.1550
			90th	0.0502	0.0100	0.0510
			95th	0.0201	0.0025	0.0250
A2	Toluene, ethylbenzene, xylenes, and styrene*	t	50th	NC	0.0625	0.1950
			75th	0.0635	0.0039	0.0500
			90th	0.0134	0.0001	0.0110
			95th	0.0033	0	0.0040
A3, B3	1,4-DCB, TCE*, PERC, chloroform, and CTC	t	50th	NC	0.0313	0.0820
			75th	0.0067	0.0010	0.0040
			90th	0.0033	0	0
			95th	0	0	0
A4	d-Limonene, α-pinene, and β-pinene	t	50th	0.3244	0.1250	0.2070
			75th	0.1171	0.0156	0.0480
			90th	0.0234	0.0010	0.0060
			95th	0.0100	0.0001	0.0030
B1	Ethylbenzene and MTBE	Gumbel	50th	0.3478	0.0625	0.3490
			75th	0.1438	0.0039	0.1430
			90th	0.0435	0.0001	0.0510
			95th	0.0234	0.0000	0.0240
B2	Benzene, MTBE, 1,4-DCB, TCE*, and PERC	t	50th	NC	0.0313	0.0630
			75th	0.0067	0.0010	0.0060
			90th	0.0033	0	0
			95th	0	0	0

Mixture ID: A: mixture identified by PMF; B: mixture identified by toxicological mode of action.
NC, not calculated as styrene and TCE had detection frequencies <50%.

Table 41. Distribution type and parameters fitted to individual VOCs (first-visit measurements) in RIOPA.

VOC	Distribution	Parameters
Benzene	Pearson5	(1.7416, 3.0237)
Toluene	Pareto	(0.80165, 3.3500)
Ethylbenzene	Lognormal	(2.3804, 3.4359)
Xylenes	Loglogistic	(0.74464, 4.8664, 1.5276)
MTBE	LogLogistic	(-0.068879, 6.9726, 1.5498)
Styrene	Pearson5	(1.4394, 0.62596)
1,4-DCB	Lognormal	(51.195, 1100.2)
TCE	Pareto	(1.0292, 0.12000)
PERC	Loglogistic	(-134.65, 136.13, 55.589)
Chloroform	Pearson5	(1.1756, 0.92852)
CTC	Loglogistic	(-0.089049, 0.70987, 5.0349)
d-Limonene	Pearson5	(1.2177, 11.984)
α-Pinene	Pearson5	(0.80312, 0.93957)
β-Pinene	Pareto	(0.77374, 0.50500)

Parameters for Pearson 5 are α, β; parameters for Pareto are θ, a; parameters for lognormal are μ, σ; parameters for loglogistic are γ, β, α.

Table 42. Goodness-of-fit statistics of fitted copulas for RIOPA mixtures.

Mixture ID	Copula	-BIC	-AIC
A1	Gaussian	113.66	117.34
	t	117.67	125.03
	Gumbel	**123.99**	**131.35**
	Clayton	112.76	120.12
	Frank	102.30	109.66
A2	Gaussian	607.97	629.88
	t	**655.80**	**681.32**
	Gumbel	327.11	330.80
	Clayton	227.24	230.93
	Frank	381.40	385.09
A3, B3	Gaussian	77.67	113.91
	t	**86.12**	**125.91**
	Gumbel	59.92	63.60
	Clayton	44.30	47.98
	Frank	54.34	58.03
A4	Gaussian	281.49	292.51
	t	**319.30**	**333.96**
	Gumbel	310.60	314.28
	Clayton	264.28	267.97
	Frank	321.34	325.02
B1	Gaussian	83.59	87.27
	t	94.59	101.95
	Gumbel	**99.17**	**106.53**
	Clayton	94.78	102.14
	Frank	81.80	89.16
B2	Gaussian	140.72	176.97
	t	**156.22**	**196.01**
	Gumbel	36.37	40.05
	Clayton	33.11	36.80
	Frank	27.53	31.21

BIC, Bayesian information criterion; AIC, Akaike information criterion.
The lowest value of the information criterion was the best-fit copula, which was shown in bold type.

Table 43. Parameters and correlation matrixes of the fitted copulas for VOC mixtures in RIOPA.

Mixture ID	Parameter
A1	θ=1.67
B1	θ=1.57

Mixture A2 (df = 4)

	Toluene	Ethylbenzene	Xylenes	Styrene
Toluene	1.00	0.64	0.65	0.12
Ethylbenzene	0.64	1.00	1.00	0.17
Xylenes	0.65	1.00	1.00	0.17
Styrene	0.12	0.17	0.17	1.00

Mixture A3 (df = 5)

	1,4-DCB	TCE	PERC	Chloroform	CTC
Benzene	1.000	-0.022	-0.015	0.011	0.004
MTBE	-0.022	1.000	0.849	0.069	0.147
1,4-DCB	-0.015	0.849	1.000	0.033	0.031
TCE	0.011	0.069	0.033	1.000	0.748
PERC	0.004	0.147	0.031	0.748	1.000

Mixture A4 (df = 2)

	d-Limonene	α-Pinene	β-Pinene
d-Limonene	1.00	0.39	0.17
α-Pinene	0.39	1.00	0.42
β-Pinene	0.17	0.42	1.00

Mixture B2 (df = 5)

	Benzene	MTBE	1,4-DCB	TCE	PERC
Benzene	1.000	0.471	0.054	0.046	0.017
MTBE	0.471	1.000	0.034	-0.010	-0.006
1,4-DCB	0.054	0.034	1.000	-0.022	-0.015
TCE	0.046	-0.010	-0.022	1.000	0.849
PERC	0.017	-0.006	-0.015	0.849	1.000

df, degree of freedom.

Table 44. Median mixture fractions based on observations and copula simulations in RIOPA.

Mixture ID	VOCs	Mixture fractions* for indicated percentile							
		Observed (n = 299)				Best-fit copula (n = 1,000)			
		50 - 75th	75 - 90th	90 - 95th	95 - 100th	50 - 75th	75 - 90th	90 - 95th	95 - 100th
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
	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

Mixture ID: A indicates mixtures indentified by PMF; B indicates mixtures identified by toxicological mode of action.
* median fractions. They may not sum to 1.
Dominant mixture fraction shown in bold.
Copula simulations use fitted marginal distributions shown in Table 41, and best-fit copula type in Table 42.

Table 45. Comparison of mixture fractions for mixture A3/B3 and B2 in RIOPA for different copulas types.

Copula	Components	Median fractions* at different percentiles of cumulative exposure			
		50th-75th	75th-90th	90th-95th	95th-100th
t	1,4-DCB	**0.447**	**0.786**	**0.968**	**0.994**
	TCE	0.031	0.010	0.002	0.000
	PERC	0.128	0.031	0.009	0.001
	Chloroform	0.134	0.052	0.013	0.001
	CTC	0.069	0.024	0.006	0.001
	Benzene	0.093	0.068	0.022	0.004
	MTBE	**0.552**	**0.515**	0.159	0.023
	1,4-DCB	0.127	0.170	**0.484**	**0.943**
	TCE	0.009	0.005	0.003	0.001
	PERC	0.031	0.016	0.012	0.001
Gaussian	1,4-DCB	**0.466**	**0.681**	**0.962**	**0.993**
	TCE	0.028	0.009	0.002	0.001
	PERC	0.107	0.041	0.009	0.002
	Chloroform	0.130	0.063	0.013	0.002
	CTC	0.059	0.025	0.007	0.001
	Benzene	0.092	0.065	0.040	0.013
	MTBE	**0.448**	**0.399**	**0.346**	0.063
	1,4-DCB	0.180	0.202	0.190	**0.852**
	TCE	0.010	0.005	0.003	0.001
	PERC	0.043	0.022	0.011	0.003
Gumbel	1,4-DCB	**0.449**	**0.754**	**0.937**	**0.989**
	TCE	0.026	0.011	0.003	0.001
	PERC	0.132	0.055	0.011	0.004
	Chloroform	0.131	0.055	0.025	0.003
	CTC	0.063	0.023	0.008	0.001
	Benzene	0.086	0.060	0.033	0.012
	MTBE	**0.496**	**0.396**	**0.343**	0.069
	1,4-DCB	0.163	0.189	0.332	**0.829**
	TCE	0.011	0.006	0.005	0.001
	PERC	0.043	0.023	0.015	0.007
Clayton	1,4-DCB	**0.418**	**0.774**	**0.946**	**0.990**
	TCE	0.025	0.010	0.003	0.001
	PERC	0.123	0.040	0.013	0.002
	Chloroform	0.134	0.051	0.013	0.002
	CTC	0.056	0.021	0.007	0.002
	Benzene	0.089	0.047	0.028	0.006
	MTBE	**0.425**	**0.439**	0.128	0.045
	1,4-DCB	0.226	0.237	**0.699**	**0.906**
	TCE	0.010	0.005	0.003	0.001
	PERC	0.040	0.026	0.013	0.005
Frank	1,4-DCB	**0.402**	**0.663**	**0.928**	**0.991**
	TCE	0.027	0.008	0.003	0.000
	PERC	0.120	0.046	0.012	0.001
	Chloroform	0.130	0.080	0.019	0.003
	CTC	0.055	0.021	0.006	0.001
	Benzene	0.088	0.054	0.037	0.014
	MTBE	**0.428**	**0.361**	**0.499**	0.070
	1,4-DCB	0.160	0.229	0.199	**0.874**
	TCE	0.009	0.007	0.004	0.001
	PERC	0.041	0.030	0.019	0.006

* median fractions. They may not sum to 1.
Dominant mixture fraction shown in bold.

Table 46. Percentage of individuals exceeding individual lifetime cancer risk thresholds for VOC mixtures in RIOPA: comparison of observations, simulations using copulas, and simulations using multivariate lognormal distribution.

Mixture ID	VOC	Type	Percentage exceeding indicated cancer risks				
			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

Table 47. Fractions of personal VOCs originating indoors (at home) in RIOPA.

F_{home}	Median fraction						p-value for K-W test						
	All (n = 427-455)	CA (n = 139-156)	NJ (n = 106-128)	TX (n = 164-179)	Hot (n = 246-268)	Cool (n = 178-195)	City	Among cities in hot season	Among cities in cool season	Season	Between seasons in CA	Between seasons in NJ	Between seasons in TX
Benzene	0.72	0.64	0.76	0.73	0.70	0.74	**0.001**	0.053	**0.007**	**0.014**	0.337	0.118	**0.039**
Toluene	0.66	0.63	0.67	0.68	0.66	0.66	0.138	0.415	0.258	0.950	0.908	0.603	0.695
Ethylbenzene	0.69	0.64	0.68	0.73	0.67	0.72	**0.028**	**0.017**	0.487	0.110	**0.021**	0.614	0.772
m,p-Xylene	0.68	0.64	0.67	0.75	0.67	0.70	**0.013**	**0.041**	0.234	0.412	0.552	0.371	0.948
o-Xylene	0.69	0.65	0.67	0.71	0.69	0.68	0.072	**0.041**	0.808	0.597	0.371	0.427	0.463
MTBE	0.66	0.63	0.58	0.72	0.61	0.73	**0.004**	**0.006**	0.555	**0.001**	**0.009**	0.098	0.157
Styrene	0.74	0.72	0.79	0.72	0.75	0.72	**0.039**	**0.008**	0.780	0.068	0.377	**0.032**	0.847
1,4-DCB	0.72	0.67	0.73	0.76	0.72	0.74	0.255	0.151	0.810	0.467	0.075	0.772	0.940
TCE	0.74	0.66	0.74	0.80	0.73	0.77	**0.000**	0.052	**0.003**	0.798	0.276	0.554	0.468
PERC	0.71	0.69	0.75	0.71	0.70	0.72	0.329	0.534	0.642	0.358	0.456	0.604	0.563
Chloroform	0.74	0.74	0.70	0.81	0.74	0.74	**0.001**	**0.006**	0.138	0.280	0.439	0.404	0.921
CTC	0.75	0.72	0.74	0.79	0.76	0.75	**0.000**	**0.003**	**0.001**	0.526	0.980	**0.024**	0.454
d-Limonene	0.71	0.72	0.67	0.71	0.71	0.70	0.053	0.259	0.096	0.827	0.272	0.454	0.767
α-Pinene	0.78	0.79	0.74	0.81	0.79	0.77	**0.017**	0.063	0.220	0.629	0.840	0.920	0.423
β-Pinene	0.76	0.76	0.73	0.78	0.74	0.78	0.175	0.663	0.183	0.302	0.504	0.844	0.232

F_{home}, fraction of personal VOCs originating indoors at home; CA, Los Angeles in California; NJ, Elizabeth in New Jersey; TX, Houston in Texas; hot, hot season from May to October; cool, cool season from November to April; K-W test, Kruskal-Wallis test; n, sample size, which excluded participants with missing time fractions > 0.25 or < 0, as well as F_{home} > 1.25.
p-value < 0.05 which indicates there is evidence that at least one of the group medians is different from the others was shown in bold type.

Table 48. Fractions of personal VOCs originating outdoors (in neighborhood) in RIOPA.

$F_{outdoor}$	Median fraction						p-value for K-W test						
	All (n = 480-481)	CA (n = 164)	NJ (n = 135-136)	TX (n = 181)	Hot (n = 279-280)	Cool (n = 200-201)	City	Among cities in hot season	Among cities in cool season	Season	Between seasons in CA	Between seasons in NJ	Between seasons in TX
Benzene	0.007	0.000	0.006	0.009	0.009	0.002	**0.021**	0.406	**0.005**	**0.011**	0.053	**0.016**	0.701
Toluene	0.003	0.000	0.002	0.007	0.006	0.002	**0.017**	0.462	**0.004**	**0.008**	**0.038**	**0.032**	0.572
Ethylbenzene	0.004	0.000	0.003	0.005	0.005	0.001	0.125	0.676	**0.030**	**0.007**	0.062	**0.026**	0.476
m,p-Xylene	0.004	0.000	0.005	0.005	0.006	0.002	**0.049**	0.492	**0.022**	**0.009**	0.060	**0.047**	0.377
o-Xylene	0.004	0.000	0.005	0.006	0.006	0.001	**0.029**	0.433	**0.014**	**0.010**	0.056	**0.042**	0.395
MTBE	0.005	0.000	0.003	0.008	0.008	0.002	**0.006**	0.316	**0.002**	**0.013**	0.064	**0.013**	0.661
Styrene	0.004	0.000	0.007	0.005	0.007	0.001	0.196	0.313	0.056	**0.002**	**0.050**	**0.005**	0.490
1,4-DCB	0.000	0.000	0.000	0.000	0.001	0.000	0.409	0.816	0.072	**0.004**	0.051	**0.010**	0.509
TCE	0.006	0.000	0.003	0.012	0.010	0.002	**0.000**	**0.039**	**0.001**	**0.004**	0.061	**0.013**	0.240
PERC	0.004	0.000	0.007	0.006	0.007	0.002	**0.016**	0.274	**0.010**	**0.006**	**0.046**	**0.020**	0.527
Chloroform	0.001	0.000	0.001	0.002	0.002	0.000	0.108	0.376	**0.023**	**0.021**	0.084	**0.019**	0.970
CTC	0.010	0.000	0.011	0.017	0.014	0.004	**0.001**	0.063	**0.001**	**0.007**	0.059	**0.004**	0.618
d-Limonene	0.000	0.000	0.000	0.000	0.000	0.000	0.721	0.145	0.292	**0.002**	**0.012**	**0.015**	0.476
α-Pinene	0.001	0.000	0.003	0.001	0.001	0.000	0.076	**0.006**	0.532	**0.004**	**0.040**	**0.003**	0.910
β-Pinene	0.002	0.000	0.001	0.003	0.003	0.001	0.108	0.547	**0.005**	**0.018**	0.063	**0.006**	0.721

$F_{outdoor}$, fraction of personal VOCs originating outdoors in neighborhood; CA, Los Angeles in California; NJ, Elizabeth in New Jersey; TX, Houston in Texas; hot, hot season from May to October; cool, cool season from November to April; K-W test, Kruskal-Wallis test; n, sample size, which excluded participants with missing time fractions > 0.25 or < 0, as well as Foutdoor > 1.25.
p-value < 0.05 which indicates there is evidence that at least one of the group medians is different from the others was shown in bold type.

Table 49. Results of linear mixed-effect models for personal exposure to gasoline-related VOCs in RIOPA.

Variable	Group/unit	Benzene		Toluene		Ethylbenzene		m,p-Xylene		o-Xylene		MTBE		Styrene	
		β	SE	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Intercept		**2.21**	**0.41**	**3.74**	**0.37**	**1.41**	**0.42**	**2.23**	**0.37**	**0.78**	**0.29**	**1.82**	**0.32**	**1.09**	**0.33**
Visit	1	-0.03	0.07	0.12	0.09	-0.14	0.08	-0.08	0.08	-0.07	0.07	0.06	0.10	0.07	0.08
	2	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
City	Los Angeles	**-0.83**	**0.12**	0.08	0.11	**-0.37**	**0.14**	**-0.29**	**0.14**	-0.06	0.13	**-0.35**	**0.16**	**-0.23**	**0.11**
	Elizabeth	**-0.37**	**0.14**	0.06	0.13	-0.16	0.18	-0.25	0.19	-0.17	0.17	0.07	0.20	-0.11	0.10
	Houston	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
Attached garage	No	**-0.19**	**0.09**	**-0.72**	**0.25**	**-0.36**	**0.12**	**-0.36**	**0.12**	**-0.35**	**0.11**	**-0.36**	**0.12**	-0.42	0.25
Cooking	No			**0.22**	**0.09**	**0.17**	**0.08**	0.15	0.09	**0.20**	**0.08**				
Education	Less than HS	0.15	0.12												
	High school	-0.08	0.10												
	> College	Reference													
Ethnicity	White					-0.13	0.15	-0.23	0.16	-0.21	0.14				
	Mexican					0.19	0.19	0.07	0.19	0.12	0.17				
	Hispanic					0.30	0.19	0.27	0.20	**0.35**	**0.18**				
	Other					Reference		Reference		Reference					
Heating fuel	Electricity	0.20	0.18												
	Gas	**0.42**	**0.16**												
	Oil and wood	Reference													
Indoor temperature	°C	**-0.04**	**0.01**												
Inverse wind speed	$knot^{-1}$	**4.20**	**0.53**			**3.16**	**0.69**	**2.84**	**0.71**	**2.54**	**0.62**	**5.86**	**0.84**		
Log-transformed AER	hr^{-1}			**-0.30**	**0.05**	**-0.17**	**0.06**	**-0.21**	**0.06**	**-0.14**	**0.05**	-0.09	0.07		
Number of floors		**-0.15**	**0.04**									**-0.20**	**0.06**		
Number of rooms		**-0.10**	**0.03**											**-0.09**	**0.02**
Open doors or windows	No									**0.22**	**0.10**			**0.20**	**0.09**
Pumping gas	No	**-0.16**	**0.08**			**-0.24**	**0.11**	**-0.22**	**0.11**	**-0.28**	**0.10**	**-0.34**	**0.13**		
Renovation in the past year	No			**-0.30**	**0.10**										
Time spent in home	min			**-0.0002**	**0.0001**	-0.0002	0.0001							**-0.0003**	**0.0001**
Unemployed	No											0.23	0.12		
Using air cleaning devices	No					-0.27	0.18	**-0.42**	**0.18**	**-0.38**	**0.16**	-0.35	0.20		
Using nail polish remover	No			-0.29	0.17	**-0.39**	**0.16**	**-0.33**	**0.17**						
Wore powder, spray or perfume	No											**0.41**	**0.12**		

AER, air exchange rate; HS, high school.
For dichotomous variables, the reference group is "Yes"; n = 400 to 530 depending on models.
p-value < 0.05 shown in bold type.

Table 50. Results of linear mixed-effect models for personal exposure to odorant-related VOCs in RIOPA.

Variable	Group/unit	1,4-DCB		Chloroform		d-Limonene		α-Pinene		β-Pinene	
		β	SE	β	SE	β	SE	β	SE	β	SE
Intercept		**3.50**	**0.78**	**1.34**	**0.47**	**3.62**	**0.39**	**2.42**	**0.25**	**1.57**	**0.44**
Visit	1	**0.33**	**0.14**	0.15	0.09	0.10	0.15	**0.18**	**0.07**	0.08	0.10
	2	Reference		Reference		Reference		Reference		Reference	
City	Los Angeles	**-1.10**	**0.30**	**-0.45**	**0.16**	**-0.82**	**0.19**	**-0.71**	**0.13**	**-1.16**	**0.15**
	Elizabeth	**-0.81**	**0.31**	-0.06	0.17	**-1.12**	**0.22**	**-0.59**	**0.14**	**-1.06**	**0.17**
	Houston	Reference		Reference		Reference		Reference		Reference	
Air conditioning	No	**0.54**	**0.23**					**-0.51**	**0.10**	-0.20	0.13
Ambient relative humidity	%			**-0.010**	**0.005**					**-0.011**	**0.005**
Furniture refinisher in neighborhood	No	**-1.30**	**0.50**								
Waxing or polishing furniture	No	**-0.81**	**0.33**								
Keeping dogs or cats	No							0.15	0.10	**0.29**	**0.11**
Log-transformed AER	hr^{-1}			**-0.41**	**0.06**	**-0.33**	**0.08**	**-0.40**	**0.05**	**-0.31**	**0.07**
Not using fresheners or candles	No									0.32	0.18
Number of rooms		**-0.14**	**0.07**	**-0.12**	**0.04**	**-0.13**	**0.04**	**-0.10**	**0.03**		
Open doors or windows	No	**0.42**	**0.20**							0.22	0.12
Other family members took showers	No			**-0.39**	**0.15**	**-0.80**	**0.18**	**-0.41**	**0.12**	**-0.35**	**0.14**
Outdoor swimming pool or hot tub	No							**-0.31**	**0.13**		
Using heating at	< 64 °F	**0.76**	**0.26**								
	64 to 70 °F	-0.03	0.24								
	> 70 °F	Reference									
Ownership of the house	No			**0.30**	**0.14**						
Pets indoors	No			**0.32**	**0.12**						
Renovation in the past year	No					**-0.45**	**0.15**				
Restaurants or bakery in neighborhood	No	**-0.63**	**0.27**								
Unemployed	No					**-0.35**	**0.16**				
Using a clothes washer	No	**0.53**	**0.19**								
Using dishwashers	No			**-0.25**	**0.13**						
Using other heaters (non-CHS)	No					**0.55**	**0.27**				

AER, air exchange rate; CHS, central heating system.
For dichotomous variables, the reference group is "Yes"; n = 393 to 433 depending on models.
p-value < 0.05 shown in bold type.

Table 51. Results of linear mixed-effect models for personal exposure to dry-cleaning and Industrial-related VOCs in RIOPA.

Variable	Group/unit	TCE		PERC		CTC	
		β	SE	β	SE	β	SE
Intercept		**-0.79**	**0.42**	-0.48	0.49	**-0.64**	**0.23**
Visit	1	**0.18**	**0.07**	**0.19**	**0.10**	-0.01	0.03
	2	Reference		Reference		Reference	
City	Los Angeles	**0.66**	**0.14**	**0.58**	**0.18**	**-0.17**	**0.07**
	Elizabeth	**1.23**	**0.14**	**0.54**	**0.24**	-0.11	0.07
	Houston	Reference		Reference		Reference	
Ambient relative humidity	%			**-0.01**	**0.01**		
Ethnicity	White			-0.12	0.19		
	Mexican			**-0.48**	**0.23**		
	Hispanic			0.06	0.24		
	Other			Reference			
Having a fireplace	No					**-0.13**	**0.07**
Indoor temperature	°C	**-0.03**	**0.01**			0.01	0.01
Inverse wind speed	$knot^{-1}$			**4.87**	**0.83**		
Log-transformed AER	hr^{-1}			**-0.20**	**0.07**		
Not using fresheners or candles	No					**-0.20**	**0.08**
Restaurants or bakery in neighborhood	No	**0.26**	**0.13**				
Source of household water	Public	**-0.58**	**0.27**			**0.50**	**0.14**
Sweeping indoors	No			0.19	0.12		
Time spent at closed cars	min	**0.0018**	**0.0005**				
Unemployed	No			**0.42**	**0.13**		
Using air cleaning devices	No					**-0.19**	**0.08**
Vinyl, asbestos or other siding	No	**-0.25**	**0.13**				
Visited dry cleaners during past week	No			**-0.63**	**0.15**		

AER, air exchange rate.
For dichotomous variables, the reference group is "Yes"; n = 400 to 446 depending on models.
p-value < 0.05 shown in bold type.

Table 52. Results of linear mixed-effect models for indoor levels of gasoline-related VOCs in RIOPA.

Variable	Group/unit	Benzene			Toluene			Ethylbenzene			m,p-Xylene			o-Xylene			MTBE			Styrene		
		β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		2.57	0.40	**<.0001**	3.88	0.52	**<.0001**	0.58	0.48	0.22	2.65	0.49	**<.0001**	1.47	0.45	**0.00**	1.46	0.37	**0.00**	1.10	0.45	**0.02**
Visit	1	-0.22	0.08	**0.010**	0.26	0.08	**0.00**	-0.09	0.09	0.30	0.07	0.09	0.47	0.11	0.08	0.18	-0.10	0.10	0.34	0.12	0.09	0.18
	2	Reference			Reference			Reference			Reference			Reference			Reference			Reference		
City	CA	-0.52	0.12	**<.0001**	-0.18	0.13	0.16	-0.49	0.15	**0.00**	-0.22	0.16	0.17	-0.03	0.14	0.85	-0.22	0.16	0.17	-0.28	0.13	**0.03**
	NJ	-0.81	0.13	**<.0001**	-0.09	0.13	0.49	-0.30	0.19	0.12	-0.22	0.21	0.29	-0.21	0.19	0.27	0.18	0.24	0.45	-0.01	0.14	0.92
	TX	Reference			Reference			Reference			Reference			Reference			Reference			Reference		
Ambient relative humidity	%				-0.01	0.00	**0.02**	-0.01	0.00	0.09	-0.01	0.00	0.08	-0.01	0.00	**0.04**				-0.02	0.00	**<.0001**
Attached garage	No	-0.23	0.09	**0.014**				-0.38	0.12	**0.00**	-0.37	0.11	**0.00**	-0.39	0.10	**0.00**	-0.62	0.13	**<.0001**	0.28	0.12	**0.03**
Cement and other flooring	No				0.22	0.11	0.05															
Central heat	No				-0.12	0.09	0.20															
Education	Less than HS	0.34	0.13	**0.010**																		
	High school	0.03	0.10	0.735																		
	> College	Reference																				
Ethnicity	White							-0.23	0.15	0.13	-0.19	0.17	0.26	-0.14	0.15	0.35	-0.04	0.18	0.85			
	Mexican							0.19	0.18	0.31	0.19	0.20	0.34	0.24	0.18	0.19	0.53	0.21	**0.01**			
	Hispanic							0.06	0.19	0.77	0.16	0.21	0.44	0.27	0.19	0.15	0.21	0.22	0.34			
	Other							Reference			Reference			Reference			Reference					
Heating fuel	Electricity							0.47	0.22	**0.03**												
	Gas							0.48	0.20	**0.02**												
	Oil and wood							Reference														
Indoor temperature	°C	-0.04	0.01	**0.005**	-0.02	0.01	0.07															
Inverse wind speed	$knot^{-1}$							3.21	0.70	**<.0001**	3.06	0.75	**<.0001**	2.92	0.69	**<.0001**	6.13	0.82	**<.0001**	2.74	0.71	**0.00**
Logtransformed AERs	hr^{-1}				-0.34	0.05	**<.0001**				-0.19	0.06	**0.00**	-0.18	0.06	**0.00**				-0.20	0.06	**0.00**
Number of floors																	-0.13	0.05	**0.01**			
Number of rooms		-0.10	0.03	**0.000**	-0.06	0.03	**0.03**				-0.08	0.03	**0.01**	-0.09	0.03	**0.00**				-0.09	0.03	**0.00**
Open doors or windows	No										0.20	0.11	0.07	0.23	0.10	**0.03**				0.18	0.11	0.08
Professional cleaning	No	0.19	0.11	0.077				0.20	0.12	0.10												
Time spent indoors at home	min				0.00	0.00	0.18													0.00	0.00	**0.04**
Type of building	Single family home																0.15	0.15	0.30			
	Mobile home																-0.31	0.25	0.21			
	Apartment/ townhouse																Reference					
Unemployed	No	0.17	0.10	0.084				0.27	0.10	**0.01**												
Use of candles or incense	No										-0.25	0.10	**0.01**	-0.22	0.09	**0.01**						
Using air cleaning devices	No							-0.41	0.17	**0.02**	-0.61	0.19	**0.00**	-0.50	0.17	**0.00**	-0.40	0.21	0.06			

AER, air exchange rate; HS, high school.
For dichotomous variables, the reference group is "Yes"; n = 387 to 455 depending on models.
p-value < 0.05 shown in bold type.

Table 53. Results of linear mixed-effect models for indoor levels of odorant-related VOCs in RIOPA.

Variable	Group/unit	1,4-DCB			Chloroform			d-Limonene			α-Pinene			β-Pinene		
		β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		4.215	0.903	**<.0001**	-0.41	0.54	**0.45**	3.57	0.39	**<.0001**	2.48	0.21	**<.0001**	1.59	0.43	**0.00**
Visit	1	0.320	0.152	**0.037**	-0.04	0.10	0.71	-0.11	0.15	0.44	0.12	0.07	**0.09**	0.12	0.09	0.21
	2	Reference			Reference			Reference			Reference			Reference		
City	CA	-0.972	0.309	**0.002**	-0.26	0.15	**0.08**	-0.46	0.19	**0.01**	-0.47	0.12	**0.00**	-1.01	0.15	**<.0001**
	NJ	-0.612	0.413	**0.140**	-0.22	0.16	0.17	-0.53	0.29	**0.07**	-0.62	0.13	**<.0001**	-1.20	0.16	**<.0001**
	TX	Reference			Reference			Reference			Reference			Reference		
Ambient relative humidity	%				-0.01	0.00	**0.02**							-0.01	0.00	**0.01**
Ethnicity	White	-0.417	0.395	0.293				0.19	0.23	0.41						
	Mexican	0.441	0.460	0.340				0.78	0.27	**0.01**						
	Hispanic	-0.287	0.452	0.526				0.06	0.30	0.85						
	Other	Reference						Reference								
Furniture or floor was waxed or polished	No	-0.906	0.342	**0.009**												
Furniture refinisher in neighborhood	No	-1.392	0.531	**0.010**												
Indoor temperature	°C				0.05	0.02	**0.01**									
Keeping dogs or cats	No													0.35	0.11	**0.00**
Logtransformed AERs	hr^{-1}				-0.54	0.06	**<.0001**	-0.43	0.09	**<.0001**	-0.46	0.05	**<.0001**	-0.34	0.06	**<.0001**
Not using fresheners	No													0.37	0.18	**0.04**
Number of rooms		-0.141	0.073	0.055				-0.13	0.04	**0.01**	-0.07	0.03	**0.01**			
Open doors or windows	No				0.16	0.12	0.19							0.18	0.12	0.15
Other family members took showers	No				-0.40	0.13	**0.00**	-0.76	0.19	**<.0001**	-0.55	0.11	**<.0001**	-0.34	0.14	**0.02**
Outdoor temperature when heating starts	< 64 °F	0.526	0.288	0.070												
	64 to 70 °F	-0.043	0.255	0.866												
	> 70 °F	Reference														
Ownership of the house	No				0.59	0.12	**<.0001**									
Pets indoors	No				0.30	0.11	**0.01**									
Renovation in the past year	No							-0.34	0.15	**0.03**						
Spending awake time at 1st floor	Yes										-0.39	0.12	**0.00**			
Using a clothes washer	No	0.684	0.208	**0.001**												
Using central air conditioning	No							-0.44	0.17	**0.01**	-0.62	0.11	**<.0001**	-0.27	0.12	**0.03**
Using cleaning solutions	No				-0.20	0.10	**0.05**									
Using dishwashers	No				-0.34	0.12	**0.01**									
Using mothballs	No	-0.404	0.314	0.201												

AER, air exchange rate.
For dichotomous variables, the reference group is "Yes"; n = 409 to 494 depending on models.
p-value < 0.05 shown in bold type.

Table 54. Results of linear mixed-effect models for indoor levels of dry-cleaning and industrial-related VOCs in RIOPA.

Variable	Group/unit	TCE			PERC			CTC		
		β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		-0.877	0.292	**0.003**	-1.99	0.25	**<.0001**	-0.70	0.10	**<.0001**
Visit	1	0.191	0.066	**0.004**	0.05	0.09	0.59	0.05	0.05	0.30
	2	Reference			Reference			Reference		
City	CA	0.707	0.128	**<.0001**	0.98	0.17	**<.0001**	-0.06	0.05	**0.24**
	NJ	1.098	0.125	**<.0001**	1.20	0.16	**<.0001**	-0.11	0.06	0.08
	TX	Reference			Reference			Reference		
Cooking	No				0.20	0.09	**0.03**			
Having a fireplace	No							0.11	0.05	**0.04**
Inverse wind speed	$knot^{-1}$				4.00	0.78	**<.0001**			
Logtransformed AERs	hr^{-1}	-0.17	0.05	**0.001**	-0.30	0.06	**<.0001**			
Professional cleaning	No				-0.28	0.13	**0.03**			
Source of household water	Public	-0.49	0.23	**0.039**						
Sweeping indoors	No				0.16	0.10	0.13			
Unemployed	No				0.24	0.12	**0.04**			
Using central air conditioning	No							-0.11	0.05	**0.03**
Using other heaters	No	-0.34	0.14	**0.020**				0.15	0.08	0.07
Using nail polish remover	No	-0.31	0.15	**0.038**						
Vacuuming	No				0.26	0.10	**0.01**	0.12	0.04	**0.01**
Vinyl, asbestos or other siding	No	-0.22	0.11	0.052	0.38	0.13	**0.00**			
Visited dry cleaners	No				-0.34	0.14	**0.02**			

AER, air exchange rate.
For dichotomous variables, the reference group is "Yes"; n = 400 to 472 depending on models.
p-value < 0.05 shown in bold type.

Table 55. Results of linear mixed-effect models for outdoor levels of gasoline-related VOCs in RIOPA.

Variable	Group/unit	Benzene			Toluene			Ethylbenzene			m,p-Xylene			o-Xylene			MTBE			Styrene		
		β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		0.39	0.39	0.321	1.17	0.16	**<.0001**	-0.13	0.41	0.754	-0.04	0.44	0.931	-0.05	0.31	0.859	2.13	0.44	**<.0001**	-1.237	0.224	**<.0001**
Visit	1	-0.03	0.06	0.593	0.26	0.06	**<.0001**	-0.03	0.07	0.643	0.06	0.07	0.403	0.09	0.07	0.207	-0.06	0.09	0.534	-0.027	0.052	0.602
	2	Reference			Reference			Reference			Reference			Reference			Reference			Reference		
City	CA	-0.56	0.10	**<.0001**	0.06	0.09	0.517	0.01	0.10	0.947	-0.25	0.13	**0.049**	-0.03	0.11	0.765	0.02	0.14	0.874	0.558	0.079	**<.0001**
	NJ	-0.76	0.13	**<.0001**	-0.16	0.10	0.099	0.05	0.14	0.699	-0.19	0.15	0.209	-0.05	0.14	0.724	-0.09	0.18	0.624	0.735	0.108	**<.0001**
	TX	Reference			Reference			Reference			Reference			Reference			Reference			Reference		
Ambient relative humidity	%	-0.01	0.00	**<.0001**				-0.02	0.00	**<.0001**	-0.01	0.00	**<.0001**	-0.02	0.00	**<.0001**	-0.02	0.00	**0.000**	-0.01	0.00	**0.000**
Attached garage	No				-0.15	0.07	**0.042**															
Cooking	No																			-0.09	0.05	**0.047**
Crawl space	No	-0.17	0.09	**0.044**																		
Ethnicity	White	-0.20	0.11	0.069				-0.21	0.11	0.068	-0.16	0.13	0.221	-0.14	0.11	0.208	-0.14	0.15	0.337	-0.03	0.08	0.686
	Mexican	0.10	0.13	0.443				0.24	0.13	0.069	0.09	0.15	0.540	0.15	0.13	0.256	0.43	0.17	**0.016**	0.27	0.10	**0.007**
	Hispanic	-0.02	0.13	0.871				0.08	0.14	0.539	0.03	0.15	0.845	0.08	0.14	0.576	-0.02	0.18	0.927	-0.07	0.10	0.506
	Other	Reference						Reference			Reference			Reference			Reference			Reference		
Foundation of slab	No				0.15	0.06	**0.016**															
Gardening	No	0.17	0.08	**0.037**																		
House volume	m3										0.00	0.00	**0.014**	0.00	0.00	**0.040**						
Inverse wind speed	$knot^{-1}$	4.18	0.50	**<.0001**	1.91	0.49	**0.000**	4.69	0.54	**<.0001**	5.65	0.58	**<.0001**	5.50	0.54	**<.0001**	5.63	0.74	**<.0001**	3.10	0.40	**<.0001**
Near diesel vehicles	No	-0.20	0.06	**0.002**																		
No pets	No				-0.27	0.07	**0.000**															
Number of floors														-0.08	0.03	**0.031**						
Number of rooms								-0.10	0.02	**<.0001**												
Open doors or windows	No																			-0.14	0.06	**0.015**
Other family members took showers	No				-0.38	0.08	**<.0001**															
Outdoor temperature	Q1	0.39	0.09	**<.0001**	0.26	0.09	**0.003**	0.22	0.10	**0.026**	0.29	0.10	**0.005**	0.23	0.10	**0.020**	-0.03	0.13	0.839			
	Q2	0.33	0.09	**0.000**	0.27	0.08	**0.001**	0.26	0.09	**0.007**	0.22	0.10	**0.029**	0.22	0.09	**0.022**	0.24	0.13	0.069			
	Q3	-0.01	0.09	0.909	-0.10	0.08	0.247	0.08	0.09	0.378	0.02	0.10	0.868	-0.02	0.09	0.803	0.01	0.13	0.941			
	Q4	Reference			Reference			Reference			Reference			Reference			Reference					
Ownership of the house	No										0.21	0.09	**0.018**	0.23	0.08	**0.005**						
Pets indoors	No										0.14	0.08	0.074	0.15	0.07	**0.033**						
Professional cleaning	No				0.14	0.08	0.079															
Tobacco products smoked in home	No	0.66	0.26	**0.012**				0.64	0.27	**0.019**	0.79	0.29	**0.008**									
Type of building	Single family home																			-0.08	0.06	0.210
	Mobile home																			0.19	0.10	0.056
	Apartment/ townhouse																			Reference		
Unvented appliances in basement	No																-0.46	0.20	**0.025**			

For dichotomous variables, the reference group is "Yes"; n = 439 to 457 depending on models.
p-value < 0.05 shown in bold type.

Table 56. Results of linear mixed-effect models for outdoor levels of odorant-related VOCs in RIOPA.

Variable	Group/unit	1,4-DCB			Chloroform			d-Limonene			α-Pinene			β-Pinene		
		β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		-0.503	0.233	**0.032**	-1.6286	0.1561	**<.0001**	-0.05043	0.2266	0.824	-1.2703	0.18	**<.0001**	0.2626	0.1449	0.071
Visit	1	0.026	0.090	0.775	0.05289	0.05428	0.331	0.1084	0.08358	0.197	0.01116	0.07082	0.875	-0.00363	0.03871	0.925
	2	Reference			Reference			Reference			Reference			Reference		
City	CA	0.798	0.116	**<.0001**	0.4407	0.08141	**<.0001**	0.9777	0.1205	**<.0001**	1.4284	0.1013	**<.0001**	-0.5964	0.04991	**<.0001**
	NJ	0.632	0.151	**<.0001**	0.3717	0.1102	**0.001**	0.5528	0.143	**0.000**	1.3766	0.1137	**<.0001**	-0.7679	0.06444	**<.0001**
	TX	Reference			Reference			Reference			Reference			Reference		
Air conditioning	No	0.35	0.10	**0.001**												
Ambient relative humidity	%													0.00	0.00	**0.020**
Attached garage	No				-0.20	0.07	**0.005**									
Cement and other flooring	No				0.16	0.07	**0.014**									
Detached garage or carport	No													0.08	0.04	**0.039**
Ethnicity	White				0.12	0.09	0.152									
	Mexican				0.27	0.10	**0.008**									
	Hispanic				0.08	0.11	0.447									
	Other				Reference											
Furniture refinisher in neighborhood	No	-0.77	0.21	**0.000**												
Inverse wind speed	$knot^{-1}$										1.56	0.54	**0.005**			
No pets	No				-0.17	0.06	**0.009**	-0.20	0.10	0.059				-0.09	0.04	**0.044**
Not using fresheners	No				-0.17	0.09	0.056	-0.28	0.14	**0.048**						
Number of floors					0.08	0.02	**0.001**	0.09	0.04	**0.033**						
Outdoor temperature	Q1	-0.36	0.13	**0.005**	-0.21	0.07	**0.005**	-0.47	0.12	**<.0001**	-0.23	0.10	**0.016**			
	Q2	-0.32	0.12	**0.011**	-0.17	0.07	**0.019**	-0.52	0.11	**<.0001**	-0.29	0.09	**0.002**			
	Q3	0.03	0.12	0.804	-0.10	0.07	0.146	-0.21	0.11	0.067	-0.13	0.09	0.146			
	Q4	Reference			Reference			Reference			Reference					
Outdoor temperature	°C													0.01	0.00	**0.033**
Professional cleaning	No							-0.31	0.11	**0.008**	-0.20	0.09	**0.024**			
Type of building	Single family home										-0.12	0.08	0.136			
	Mobile home										0.11	0.12	0.365			
	Apartment/ townhouse										Reference					
Using mothballs	No	0.15	0.09	0.117												
Using cloth dryers	No										0.13	0.07	0.055			
Wore any powder/hair spray/perfume	No													0.07	0.04	0.062

For dichotomous variables, the reference group is "Yes"; n = 437 to 470 depending on models.
p-value < 0.05 shown in bold type.

Table 57. Results of linear mixed-effect models for indoor levels of dry-cleaning and industrial-related VOCs in RIOPA.

Variable	Group/unit	TCE			PERC			CTC		
		β	SE	p-value	β	SE	p-value	β	SE	p-value
Intercept		-1.86	0.14	**<.0001**	-2.26	0.20	**<.0001**	-0.29	0.08	**0.00**
Visit	1.00	0.14	0.04	**0.00**	-0.03	0.08	0.70	0.06	0.03	0.08
	2.00	Reference			Reference			Reference		
City	CA	0.46	0.06	**<.0001**	1.40	0.11	**<.0001**	-0.01	0.05	0.82
	NJ	0.80	0.07	**<.0001**	1.20	0.12	**<.0001**	-0.08	0.06	0.15
	TX	Reference			Reference			Reference		
Dry cleaners in neighborhood	No				-0.16	0.08	**0.05**			
Inverse wind speed	$knot^{-1}$	0.74	0.33	**0.03**	4.63	0.59	**<.0001**			
No pets	No	-0.12	0.05	**0.01**	-0.22	0.09	**0.01**			
Not using fresheners	No							-0.15	0.06	**0.01**
Number of carpeted rooms					-0.05	0.02	**0.05**			
Number of floors		-0.03	0.02	0.07						
Open doors or windows	No	-0.08	0.05	0.08				0.06	0.04	0.08
Outdoor temperature	Q1	0.13	0.06	**0.04**	0.15	0.11	0.16			
	Q2	0.17	0.06	**0.00**	0.33	0.10	**0.00**			
	Q3	0.03	0.06	0.59	0.00	0.10	0.99			
	Q4	Reference			Reference					
Type of building	Single family home							-0.12	0.04	**0.01**
	Mobile home							-0.12	0.06	0.07
	Apartment/townhouse							Reference		
Unvented appliances in basement	No	-0.23	0.09	**0.01**						
Vacuuming	No				0.16	0.07	**0.02**			

For dichotomous variables, the reference group is "Yes"; n = 402 to 461 depending on models.
p-value < 0.05 shown in bold type.

Table 58. The reduction in residual variance (R^2) attributable to fixed-effect variables in linear mixed-effect models for RIOPA VOCs.

VOCs	R^2		
	Outdoor	Indoor	Personal
Benzene	0.37	0.25	0.29
Toluene	0.23	0.09	0.10
Ethylbenzene	0.37	0.13	0.15
m,p-Xylene	0.31	0.12	0.13
o-Xylene	0.41	0.16	0.19
MTBE	0.23	0.21	0.25
Styrene	0.44	0.15	0.06
1,4-DCB	0.17	0.12	0.16
TCE	0.62	0.25	0.22
PERC	0.65	0.42	0.32
Chloroform	0.33	0.32	0.16
CTC	0.35	0.13	0.003
d-Limonene	0.29	0.27	0.26
α-Pinene	0.54	0.40	0.36
β-Pinene	0.48	0.39	0.40

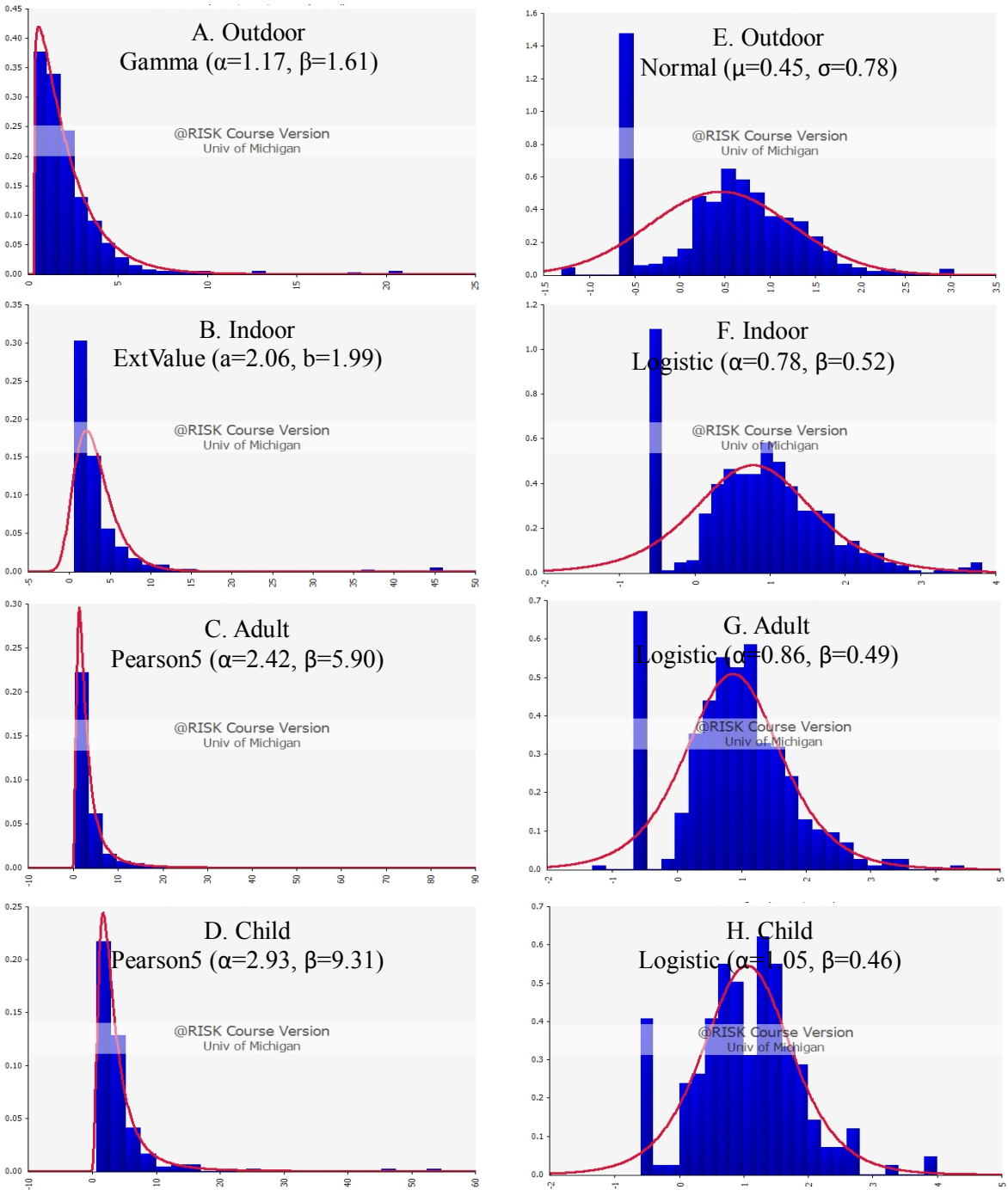

Figure 1. Observed (histograms in blue bars) and fitted distributions (red line) of benzene concentrations in RIOPA by sample type.

Left panels (A-D) are untransformed data; right panels (E-H) use natural log transform.

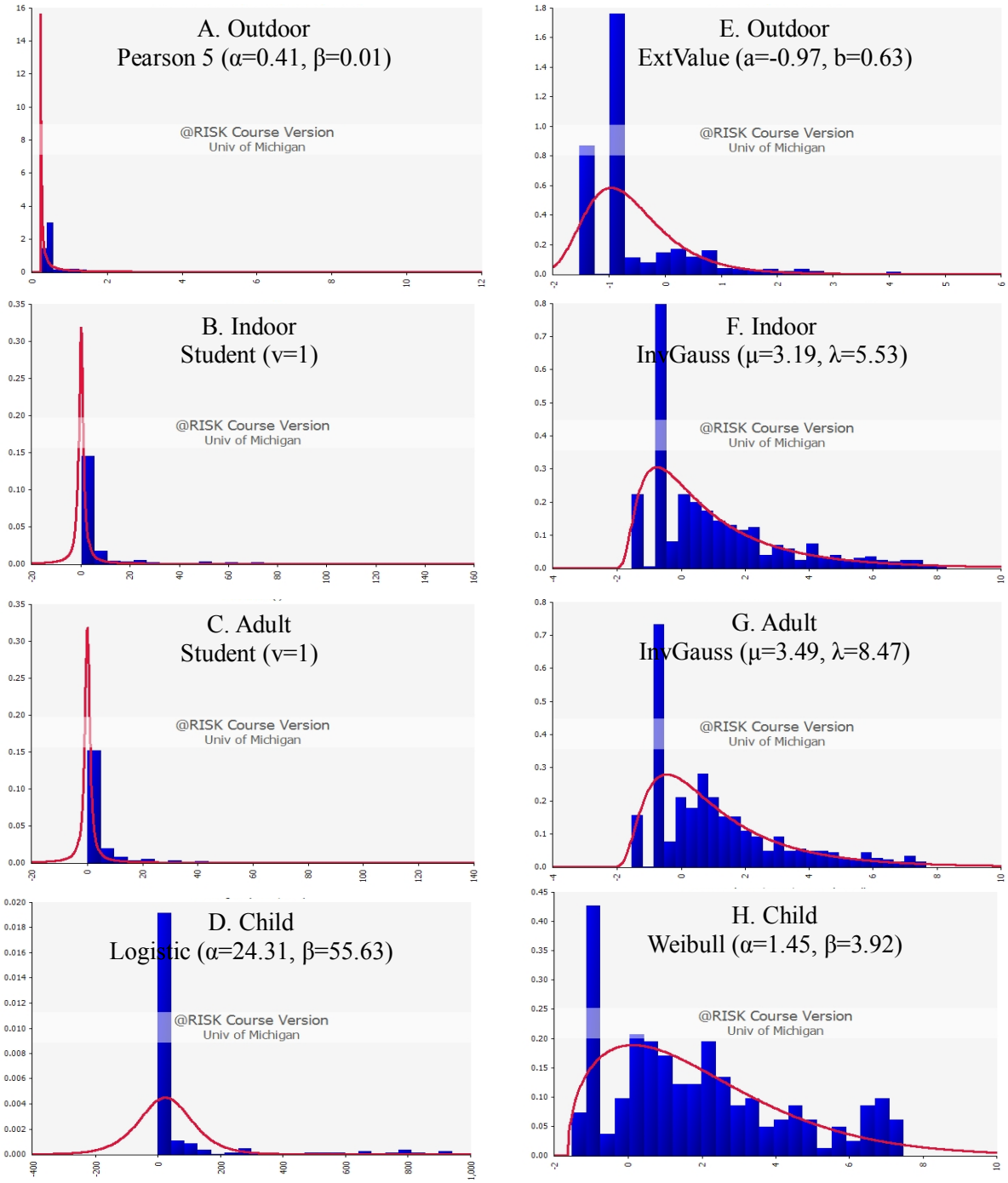

Figure 2. Observed (histograms in blue bars) and fitted distributions (red line) of 1,4-DCB concentrations in RIOPA by sample type.

Left panels (A-D) are untransformed data; right panels (E-H) use natural log transform. Plots omit the following: 1,4-DCB concentrations > 5 μg m^{-3} (n=23), 150 μg m^{-3} (n=41), 150 μg m^{-3} (n=38) and 1000 μg m^{-3} (n=10) in 2A, 2B, 2C and 2D, respectively.

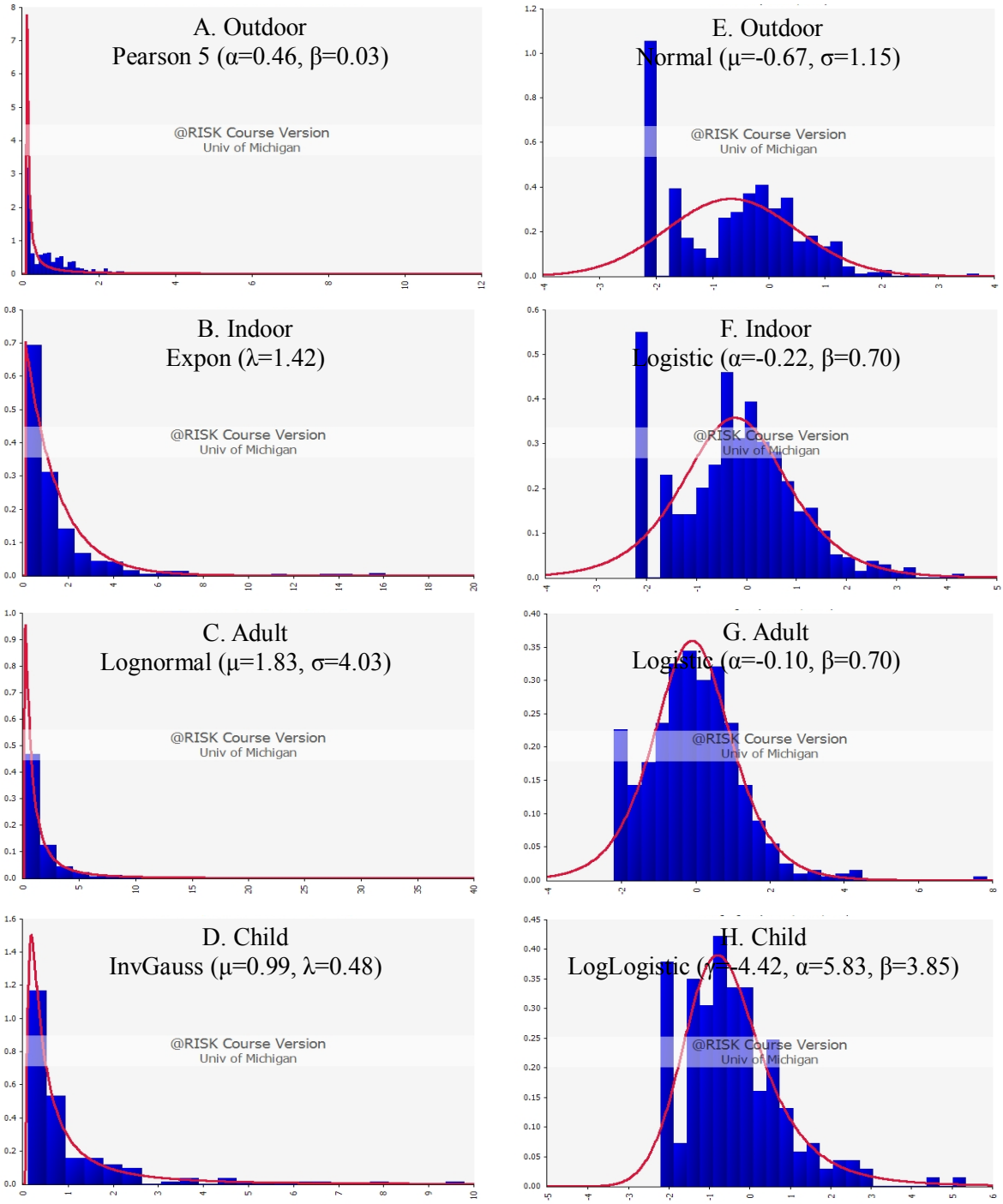

Figure 3. Observed (histograms in blue bars) and fitted distributions (red line) of PERC concentrations in RIOPA by sample type.

Left panels (A-D) are untransformed data; right panels (E-H) use natural log transform. Plots omit the following: PERC concentrations > 3 μg m^{-3} (n=32), 30 μg m^{-3} (n=1), 40 μg m^{-3} (n=6) and 20 μg m^{-3} (n=2) in 3A, 3B, 3C and 3D, respectively.

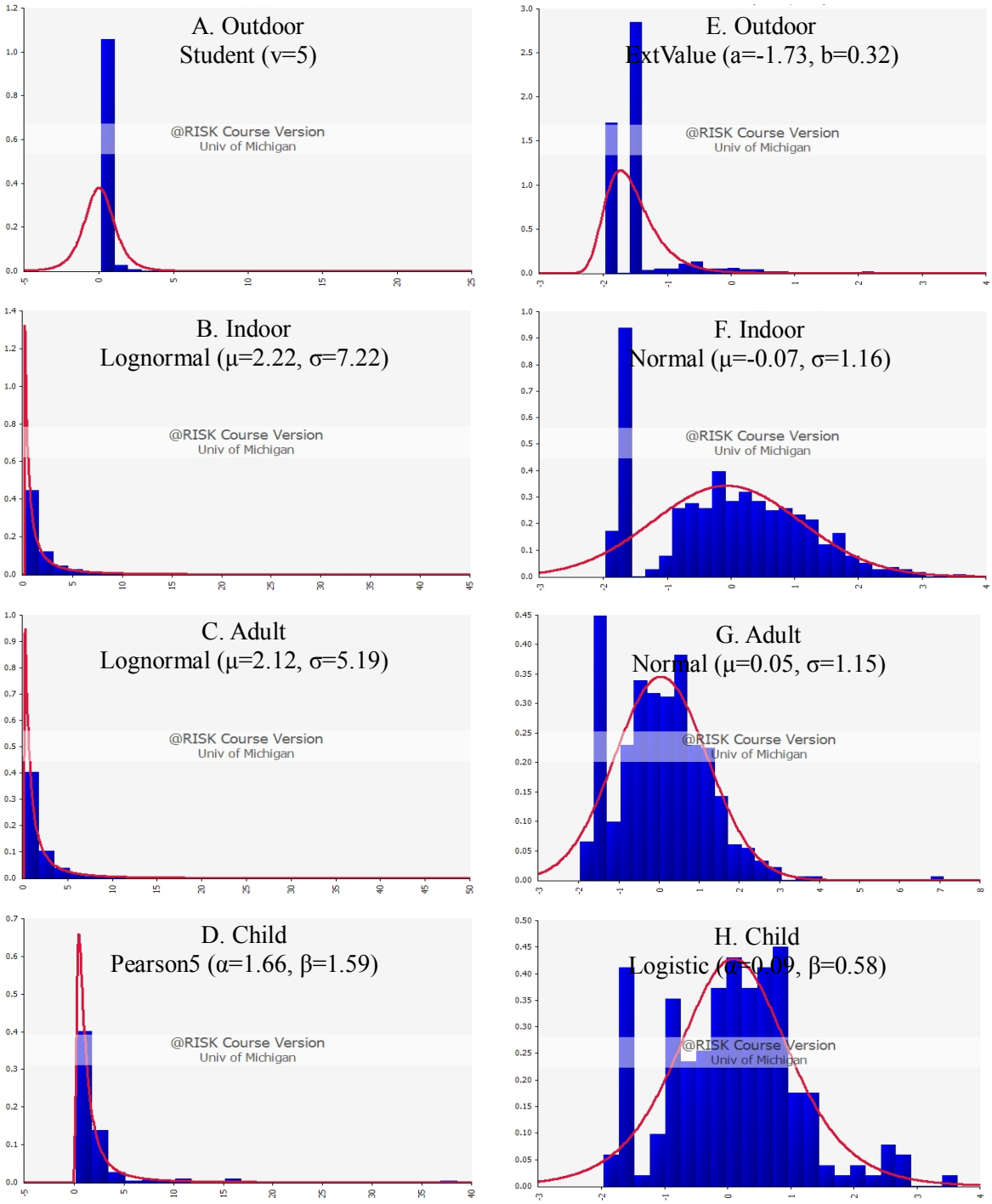

Figure 4. Observed (histograms in blue bars) and fitted distributions (red line) of chloroform concentrations in RIOPA by sample type.

Left panels (A-D) are untransformed data; right panels (E-H) use natural log transform. One extreme value of 1224 μg m^{-3} was not showed in 4C.

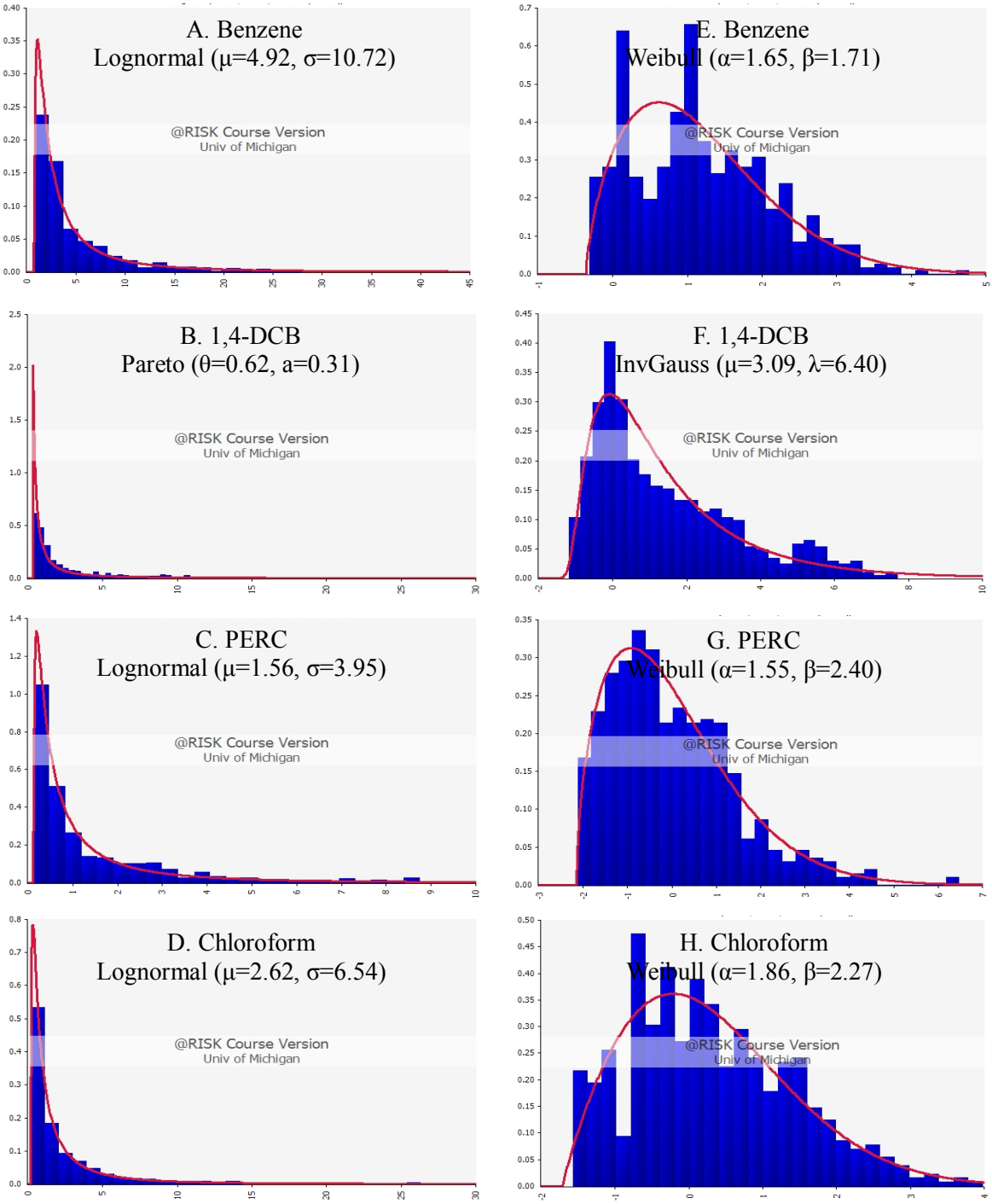

Figure 5. Observed (histograms in blue bars) and fitted distributions (red line) of benzene, 1,4-DCB, PERC and chloroform concentrations in 1999/2000 NHANES.

Left panels (A-D) are untransformed data; right panels (E-H) use natural log transform. Left plots omit the following: Benzene > 60 μg m^{-3} (n=2), 1,4-DCB >10 μg m^{-3} (n=171), PERC >10 μg m^{-3} (n=44) and chloroform >30 μg m^{-3} (n=4).

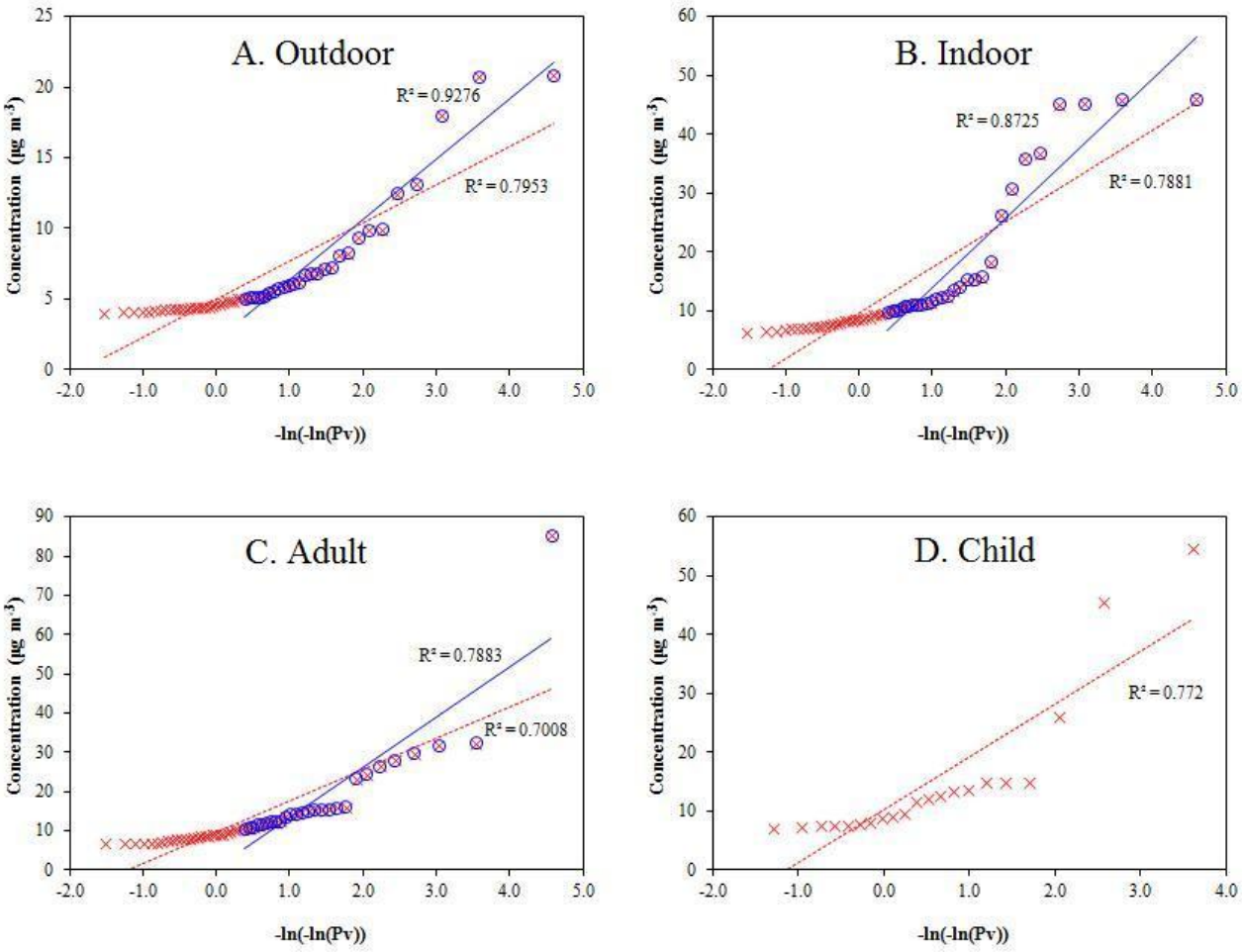

Figure 6. Top 10% (red cross and dashed line) and 5% (blue circle and solid line) of benzene concentrations in RIOPA fitted to maximum extreme distributions by sample type.

Pv = (r - 0.44)/(n + 0.12), where r = the reverse rank of Ci, and n = number of the extreme values (Barnett, 1975).

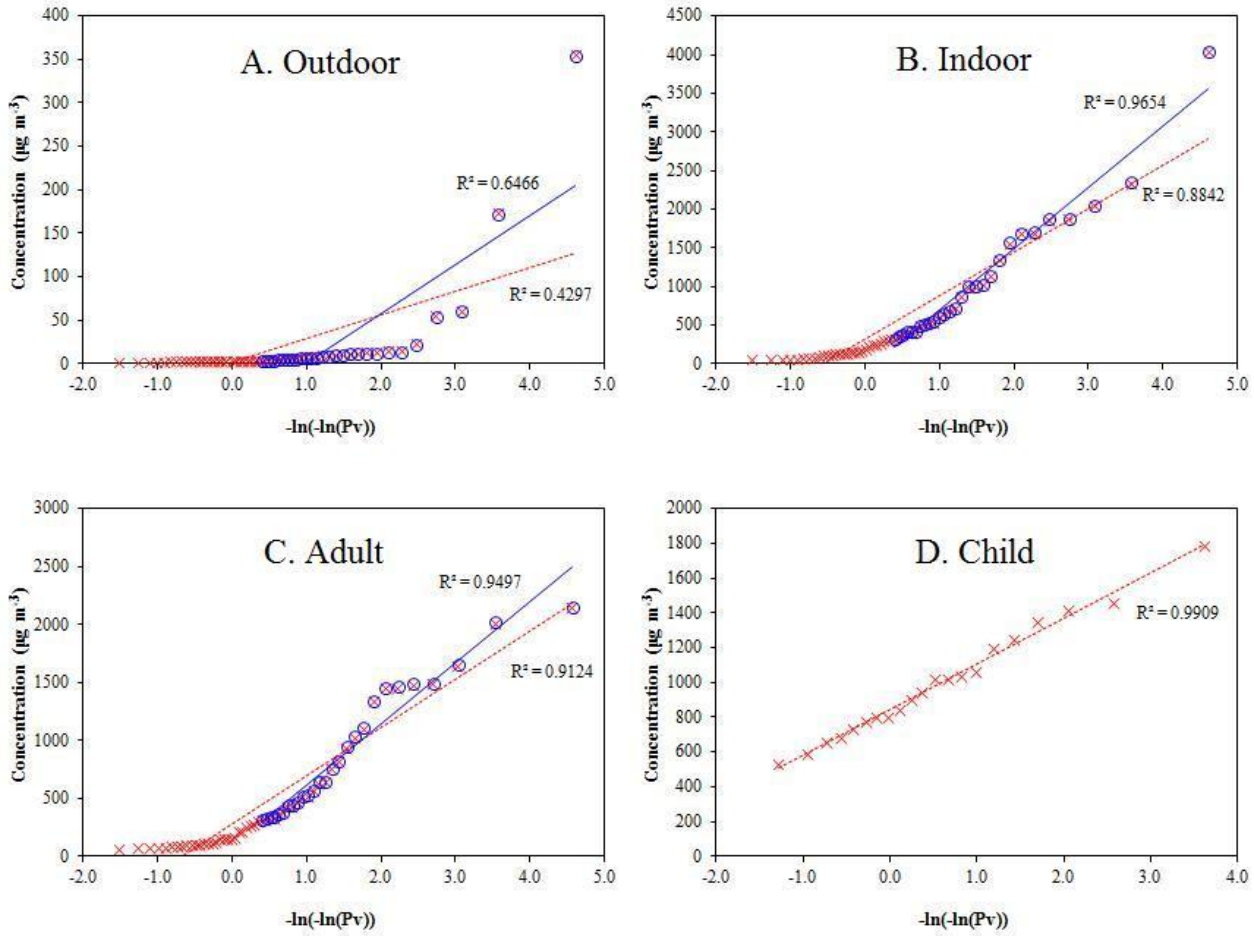

Figure 7. Top 10% (red cross and dashed line) and 5% (blue circle and solid line) of 1,4-DCB concentrations in RIOPA fitted to maximum extreme distributions by sample type.

Pv = (r - 0.44)/(n + 0.12), where r = the reverse rank of Ci, and n = number of the extreme values (Barnett, 1975).

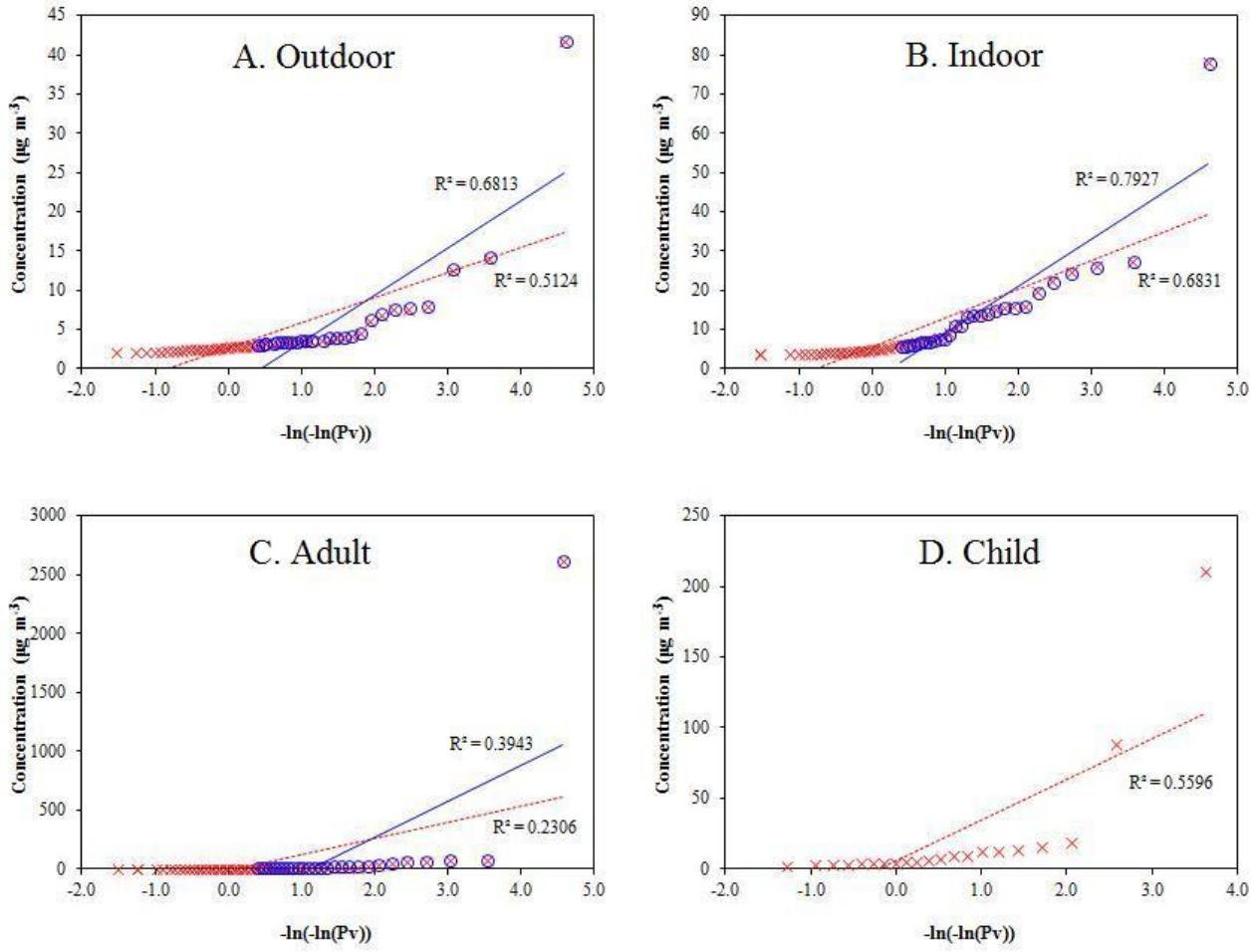

Figure 8. Top 10% (red cross and dashed line) and 5% (blue circle and solid line) of PERC concentrations in RIOPA fitted to maximum extreme distributions by sample type.

Pv = (r - 0.44)/(n + 0.12), where r = the reverse rank of Ci, and n = number of the extreme values (Barnett, 1975).

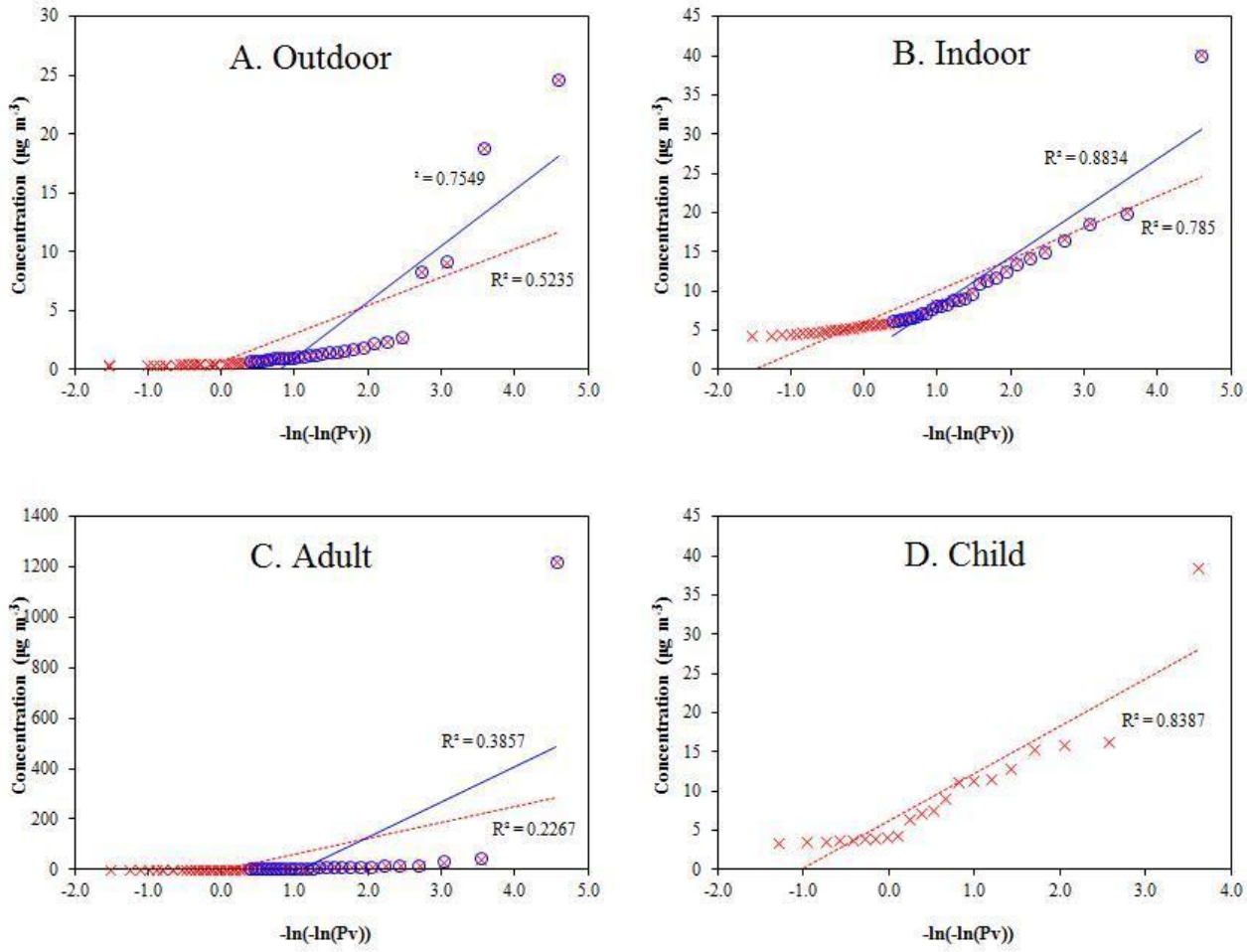

Figure 9. Top 10% (red cross and dashed line) and 5% (blue circle and solid line) of chloroform concentrations in RIOPA fitted to maximum extreme distributions by sample type.

Pv = (r - 0.44)/(n + 0.12), where r = the reverse rank of Ci, and n = number of the extreme values (Barnett, 1975).

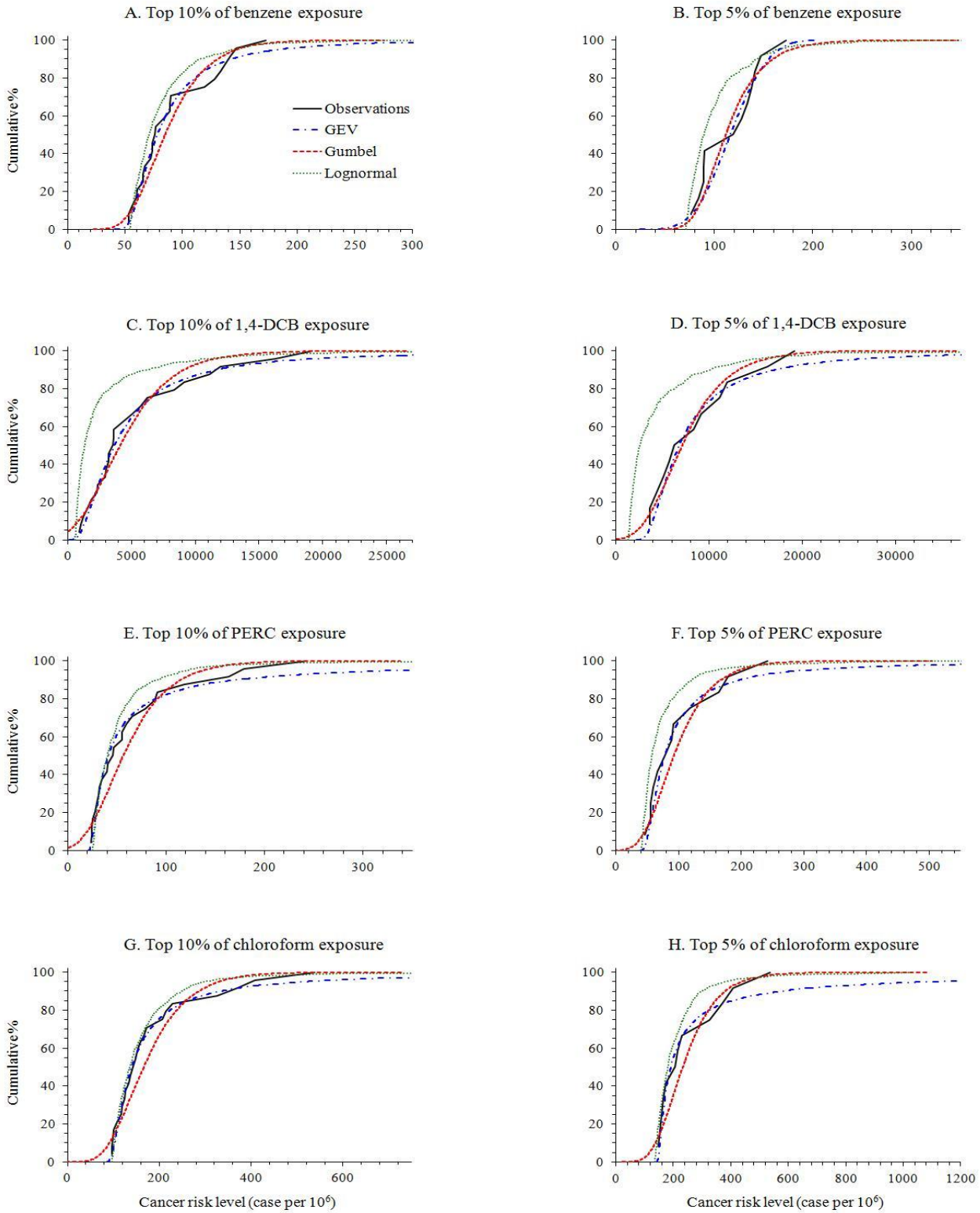

Figure 10. Comparison of cancer risks for top 10% and 5% of VOC exposure using observed measurements, generalized extreme value, Gumbel and lognormal simulations for the RIOPA data.

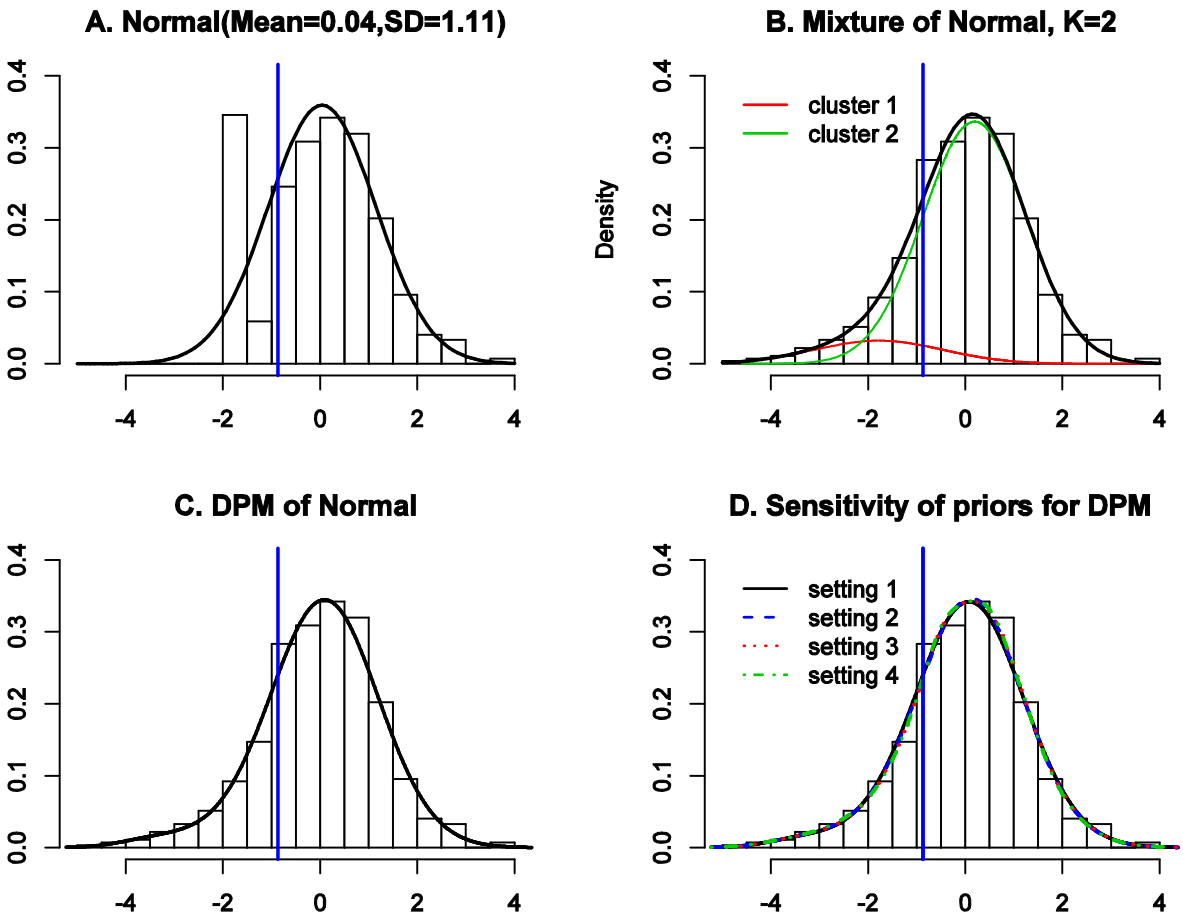

Figure 11. The fitted density plots for chloroform (log scale) in RIOPA using normal, mixture of normal and Dirichlet process mixture of normal model.

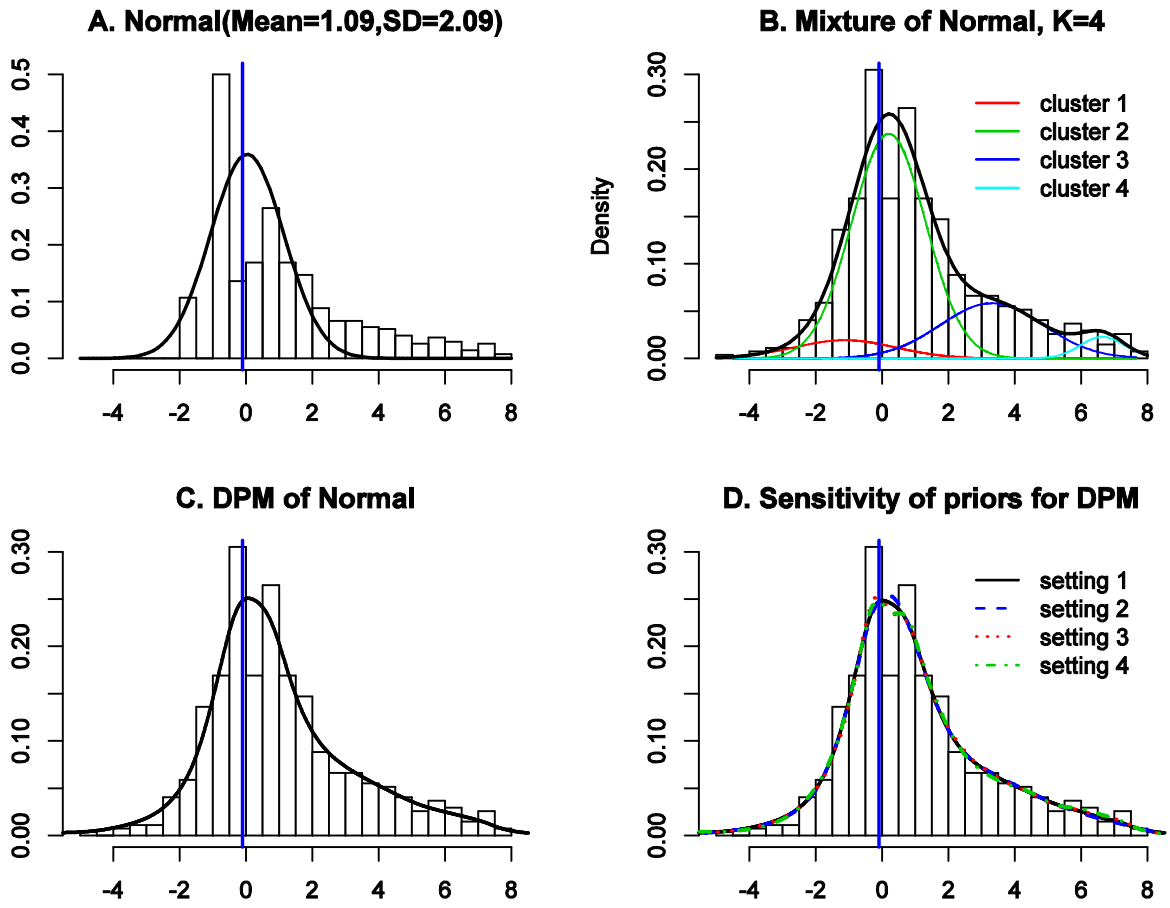

Figure 12. The fitted density plots for 1,4-DCB (log scale) in RIOPA using normal, mixture of normal and Dirichlet process mixture of normal model.

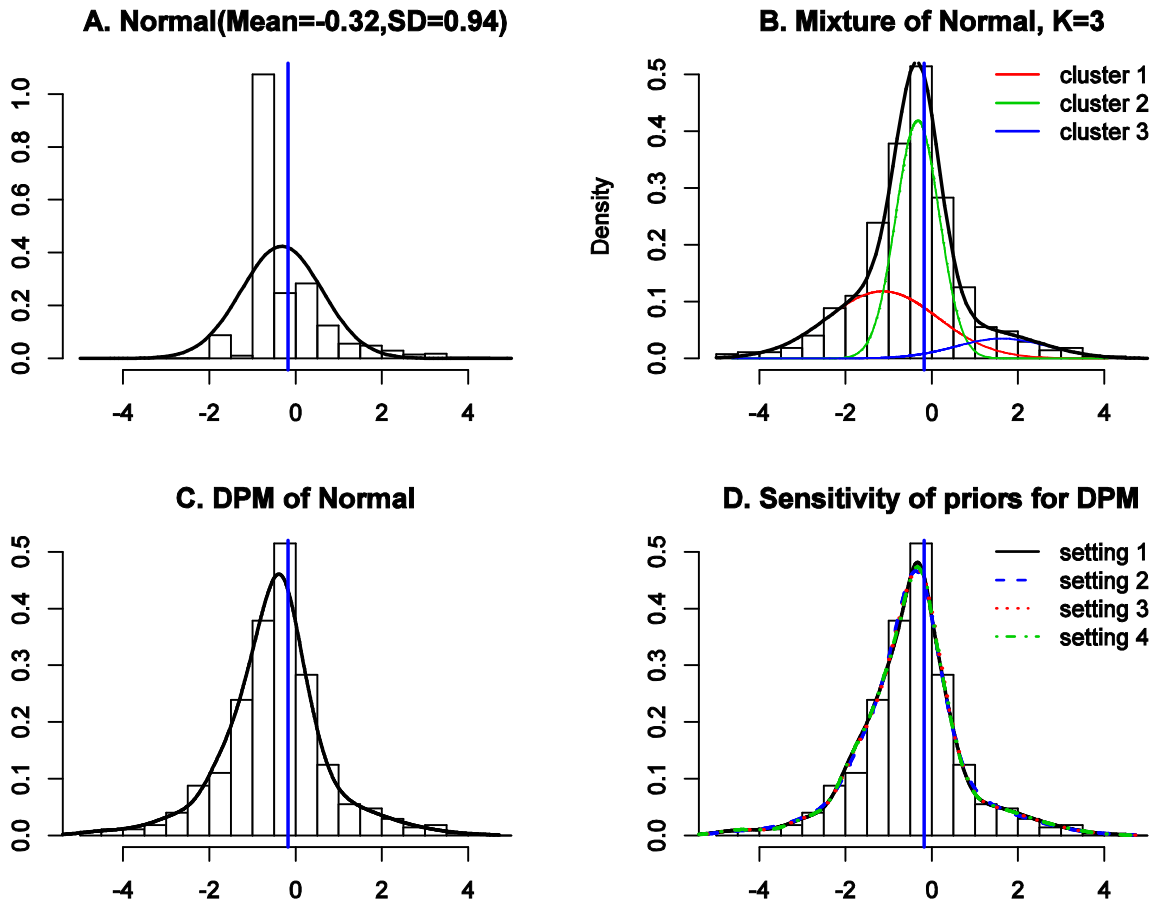

Figure 13. The fitted density plots for styrene (log scale) in RIOPA using normal, mixture of normal and Dirichlet process mixture of normal model.

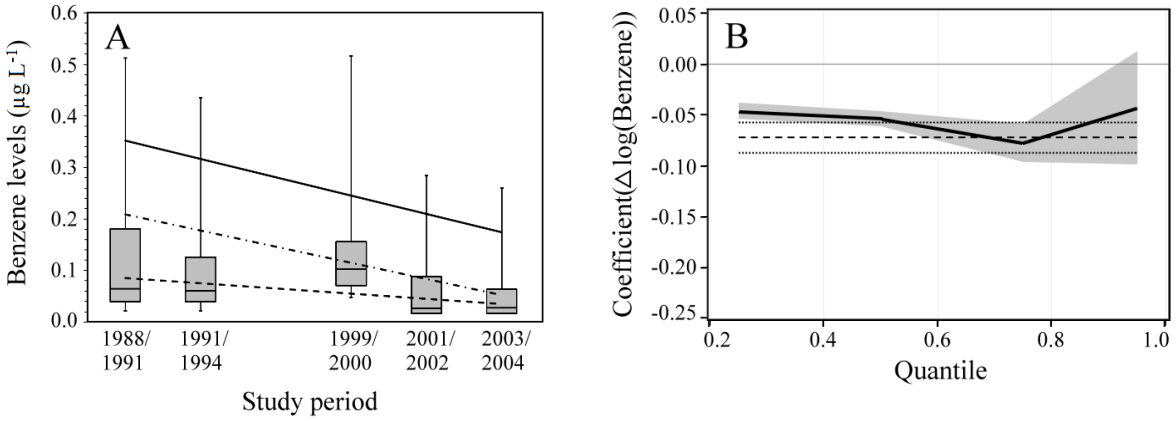

Figure 14. A) Box plot of benzene concentrations showing 0.05, 0.25, 0.50, 0.75 and 0.95 quantiles for each NHANES cohort. Linear QR trend lines for 0.5, 0.75 and 0.95 quantiles are shown as dashed, dashed and dotted, and solid lines, respectively.

B) Quantile plot for linear QR model of benzene over entire study period (1988-2004). A solid line shows coefficients for linear QR models at various quantiles. A dashed horizontal line shows coefficients for linear regression model, and dotted horizontal lines show 95% confidence intervals.

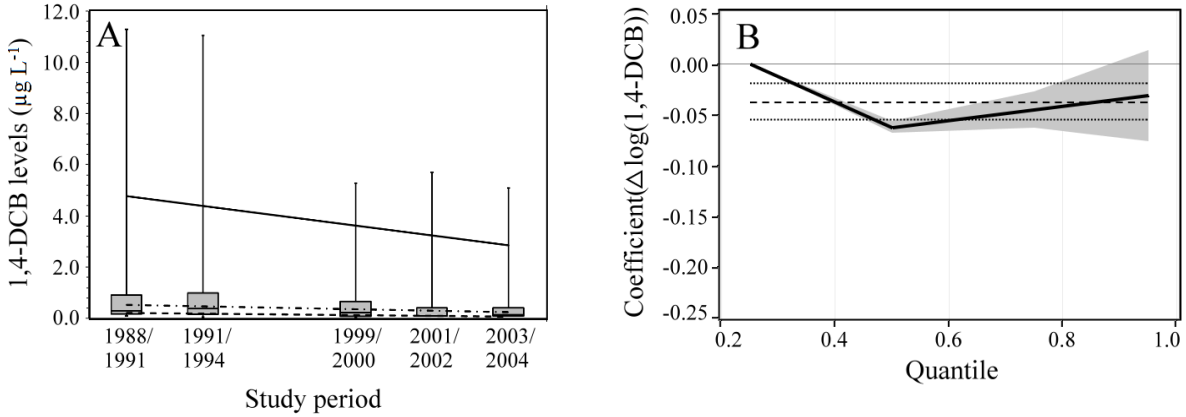

Figure 15. Box plots and linear QR model results for 1,4-DCB. Otherwise as Figure 1.

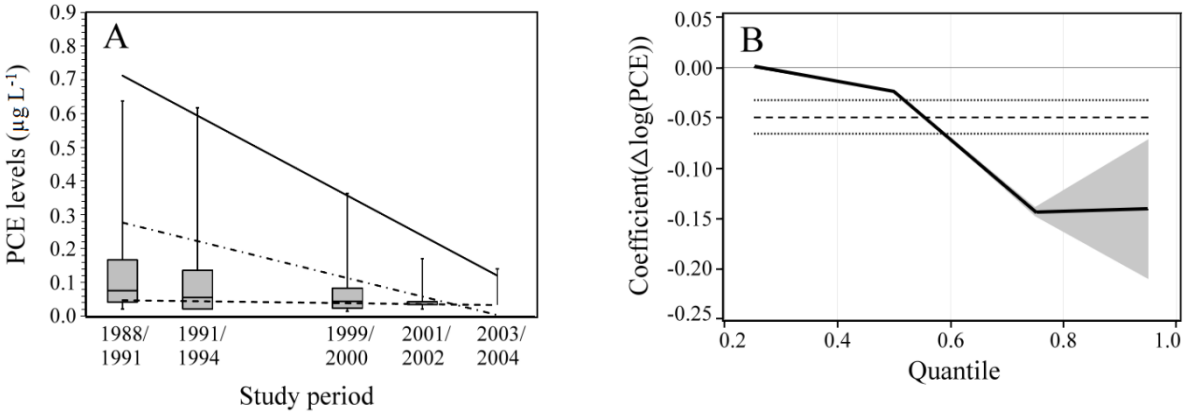

Figure 16. Box plots and linear QR model results for PERC. Otherwise as Figure 1.

Figure 17. Outdoor VOC composition at quintiles of total VOC concentrations in RIOPA. Warm season indicates April to September, and cold season indicates October to March. , aromatics; , MTBE; , chlorocarbons; , terpenes.

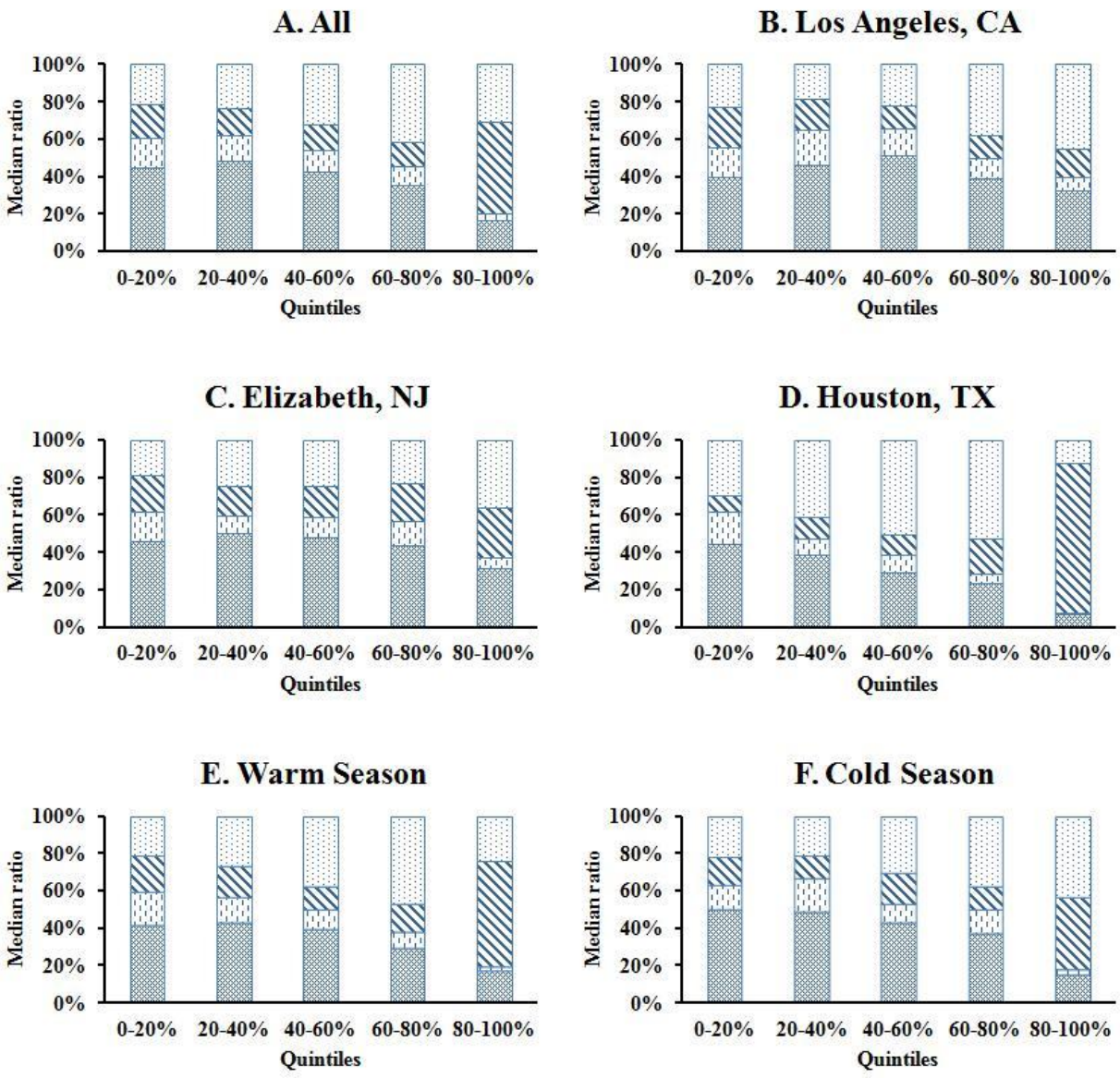

Figure 18. Indoor VOC composition at quintiles of total VOC concentrations in RIOPA. Warm season indicates April to September, and cold season indicates October to March. , aromatics; , MTBE; , chlorocarbons; , terpenes.

Figure 19. Personal VOC composition at quintiles of total VOC concentrations in RIOPA. Warm season indicates April to September, and cold season indicates October to March. , aromatics; , MTBE; , chlorocarbons; , terpenes.

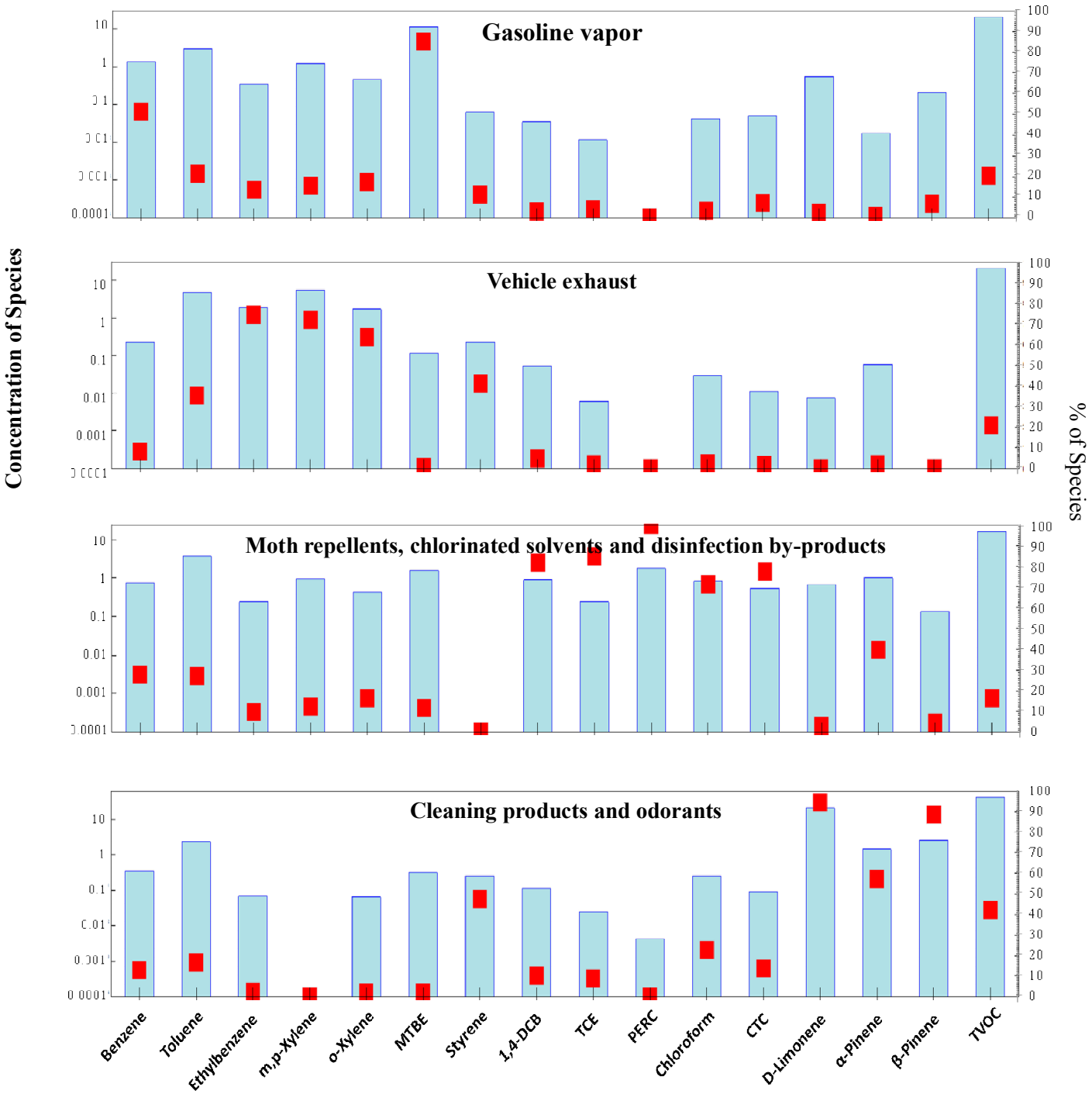

Figure 20. Factor profiles from PMF analyses for personal exposure measurements of VOCs in RIOPA.

Red boxes indicate percentage of mass of each species apportioned to the factor; blue bars indicate concentrations of species.

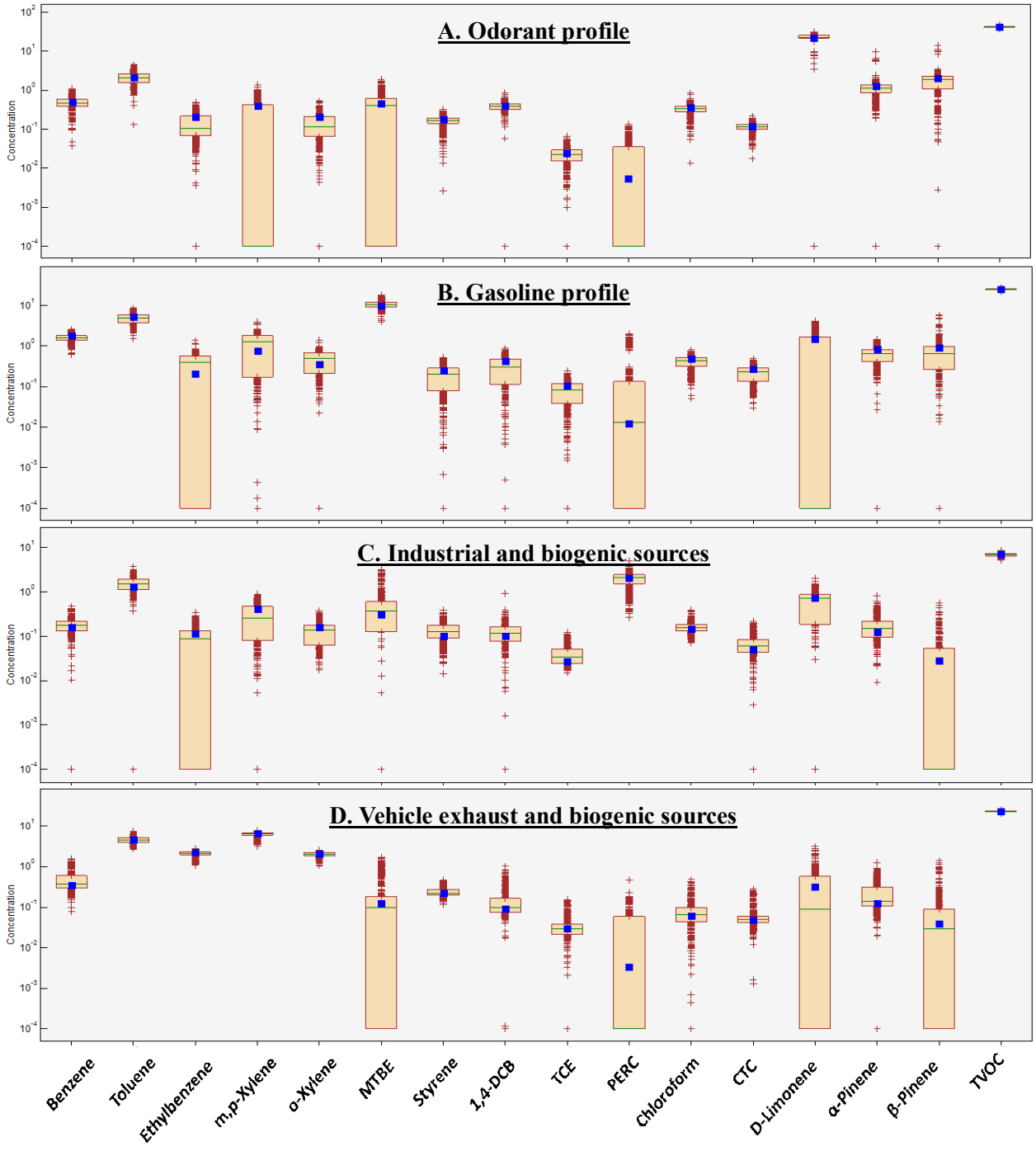

Figure 21. Factor profiles and variability for personal VOC exposures in RIOPA based on bootstrap analyses.

Blue boxes show original factor profiles; red boxes show interquartile ranges; green lines are the medians of the bootstrap results; and red crosses are values outside the interquartile range.

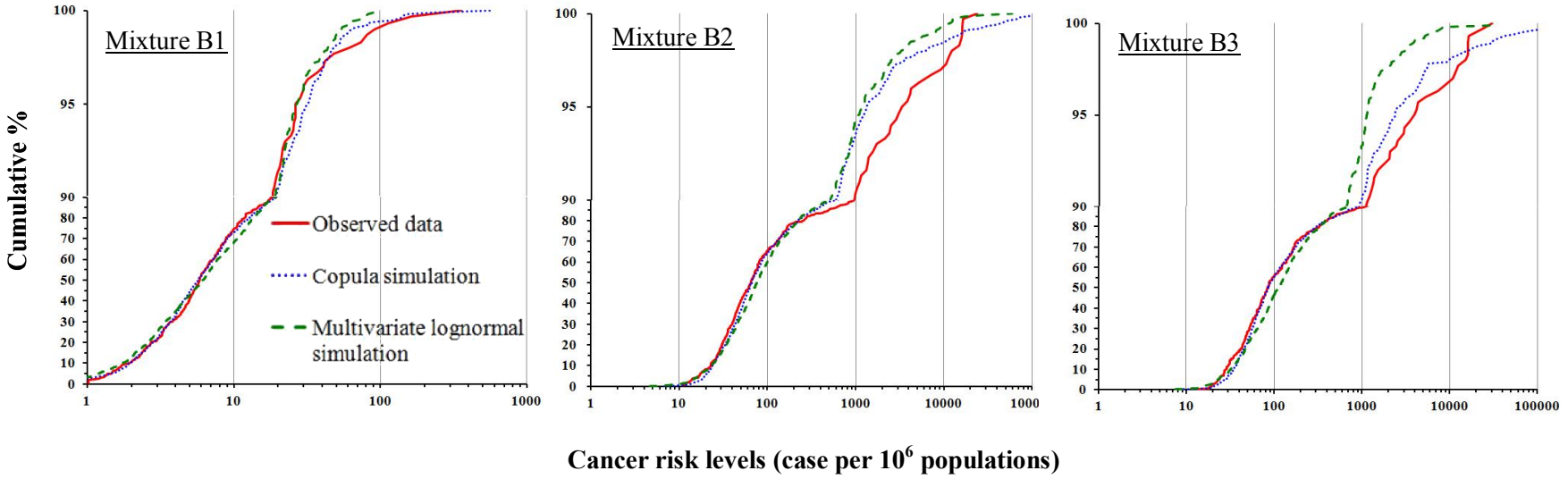

Figure 22. Cumulative probability plots of cancer risks for VOC mixtures using observations, copula and multivariate lognormal simulations in the RIOPA study.

The y-axis scale emphasizes differences at upper percentiles.

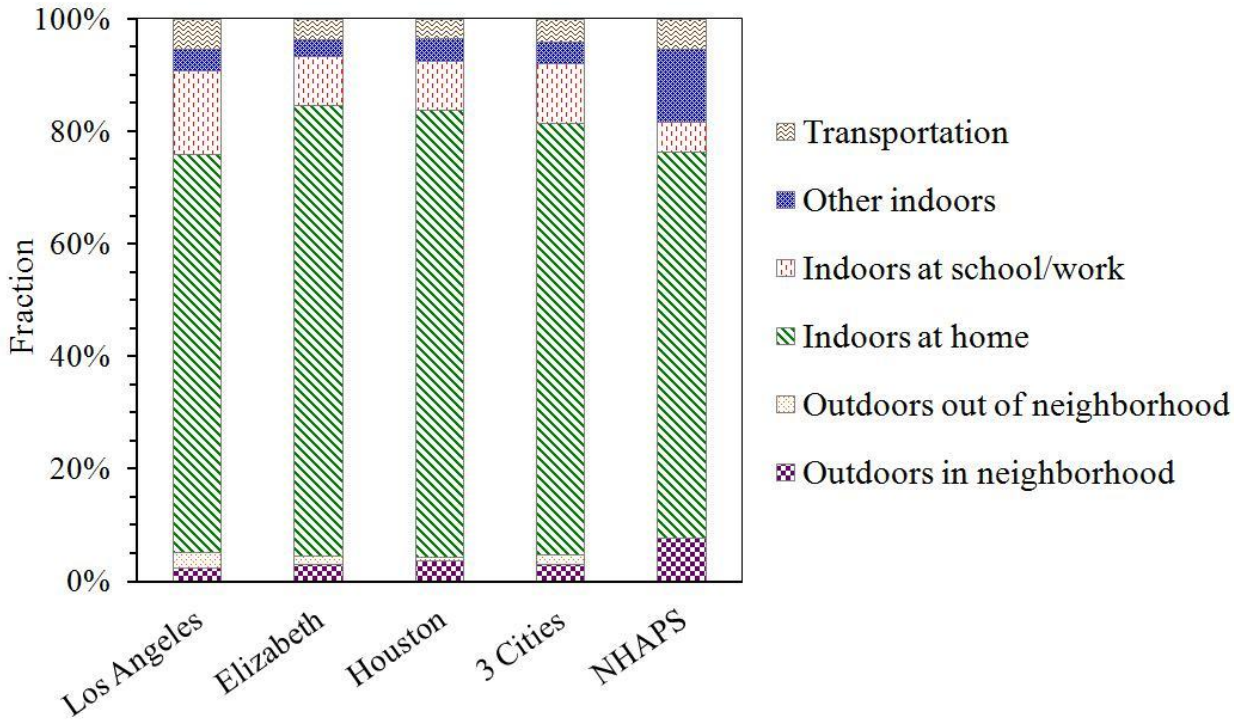

Figure 23. Mean time-spent fractions for RIOPA (by city) and NHAPS participants (Klepeis et al. 2001).

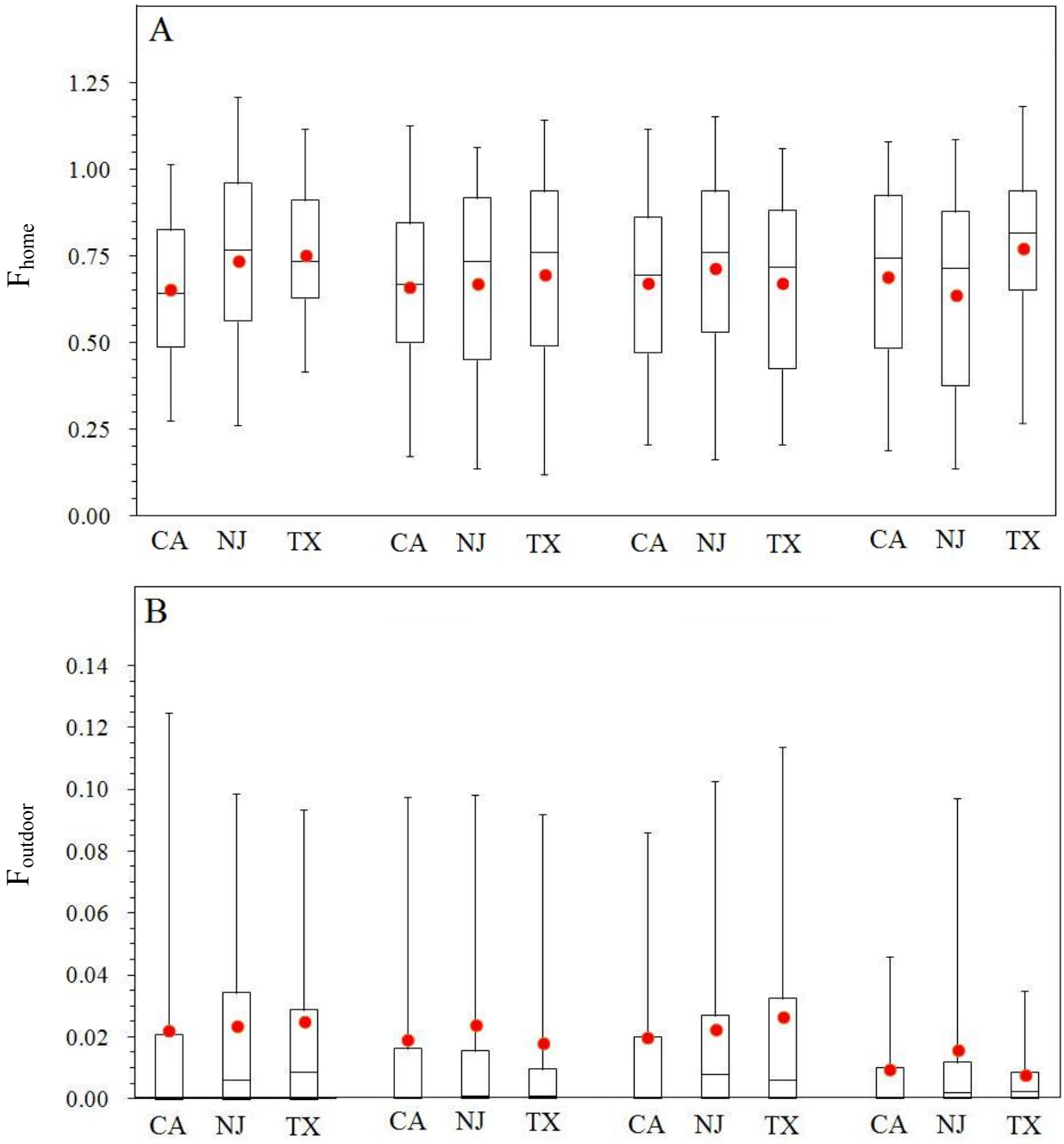

Figure 24. Box plots showing 5th, 25th, 50th, 75th and 95th percentiles and average (red dot) for F_{home} and $F_{outdoor}$ for selected VOCs in the three RIOPA cities.

CA, Los Angeles; NJ, Elizabeth; TX, Houston.

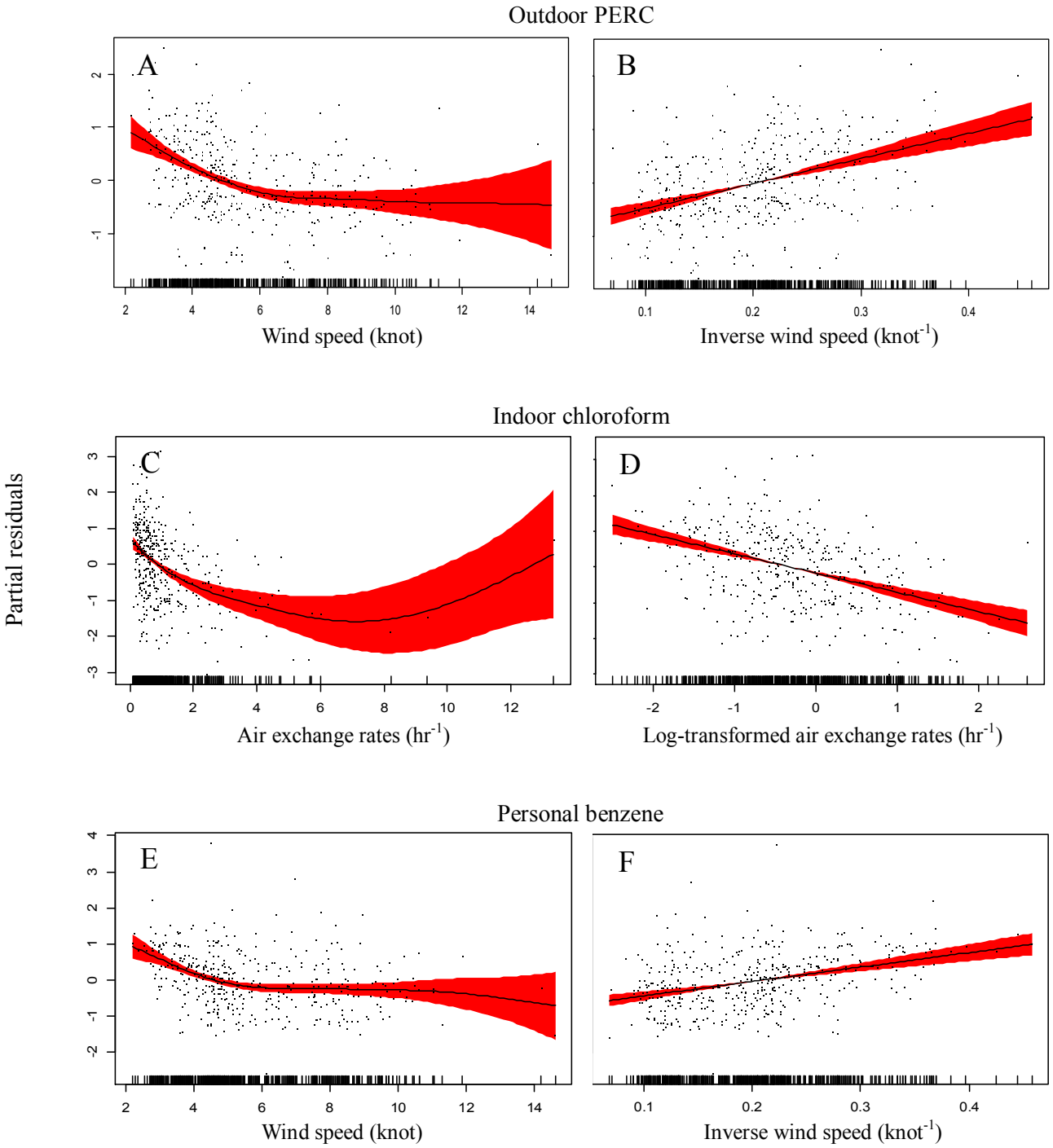

Figure 25. Partial residual plots of linear mixed-effect models for selected VOCs in RIOPA.

Supplemental Table S 1. Method detection limits and detection frequencies of VOCs concentrations in RIOPA.

VOCs	MDL (μg m^{-3})		Outdoor		Indoor		Adult		Child	
	NJ & CA	TX & CA	n	% below MDL	n	% below MDL	n	% below MDL	n	% below MDL
Benzene	1.1	0.54	555	24.5	554	17.9	544	14.5	209	8.1
CTC	0.27	0.34	555	3.2	554	4.5	544	4.6	209	3.3
Chloroform	0.42	0.28	555	87.2	554	23.3	544	17.3	209	11.5
1,4-DCB	0.91	0.43	555	71.5	554	35.1	544	28.5	209	19.6
Ethylbenzene	0.74	0.22	555	24.5	554	15.2	544	14.2	209	6.2
d-Limonene	1.27	0.74	555	79.6	554	13.4	544	11.6	209	6.2
MC	2.1	0.29	555	85	554	70.6	544	67.7	209	58.4
MTBE	0.68	0.38	555	3.6	554	6.1	544	3.9	209	3.8
α-Pinene	2.04	0.28	555	75.7	554	37.4	544	33.7	209	19.1
β-Pinene	1.01	2.09	555	93.7	554	47.8	544	43.8	209	30.6
Styrene	0.84	0.34	555	83.5	554	54.3	544	51.5	209	31.6
Toluene	6.7	7.12	555	66.1	554	30	544	25	209	22.5
TCE	0.44	0.24	555	80	554	74.2	544	68.6	209	77.5
PERC	0.42	0.22	555	31.9	554	18.6	544	12.5	209	14.8
m,p-Xylene	1.4	0.65	555	15.3	554	10.3	544	8.5	209	3.8
o-Xylene	0.85	0.29	555	26.7	554	18.2	544	12.9	209	7.2

MDL, method detection limit; n, sample size.

Supplemental Table S 2. Method detection limits and detection frequencies of VOC concentrations in blood by NHANES cohort.

VOCs	NHANES 1988-1991				NHANES 1991-1994				NHANES 1999/2000				NHANES 2001/2002				NHANES 2003/2004			
	MDL	n	Extre	DF	MDL	n	Extre	DF	MDL	n	Extre	DF	MDL	n	Extre	DF	MDL	n	Extre	DF
Aromatics																				
Benzene	0.0300	552	134	62	0.0300	466	88	71	0.0323	300	0	100	0.0170	837	0	53	0.0170	1345	0	59
Toluene	0.0920	552	362	34	0.0920	466	81	82	0.0231	304	6	97	0.0177	954	1	95	0.0177	1336	0	95
Ethylbenzene	0.0200	552	352	35	0.0200	466	60	81	0.0101	262	1	90	0.0170	879	0	61	0.0170	1299	0	68
m,p-Xylene	0.0330	552	0	37	0.0330	466	0	91	0.0358	296	2	96	0.0240	962	0	96	0.0240	1346	0	98
o-Xylene	0.0400	552	353	36	0.0400	466	37	87	0.0210	309	1	58	0.0346	981	0	40	0.0346	1365	0	37
Styrene	0.0190	552	352	34	0.0190	466	42	77	0.0066	284	1	94	0.0212	950	2	54	0.0212	1245	0	41
THMs																				
Chloroform	0.0210	552	109	45	0.0210	466	33	50	0.0064	255	2	99	0.0017	744	3	96	0.0015	1222	0	93
BDCM	0.0090	552	41	13	0.0090	466	40	12	0.0002	354	0	95	0.0002	785	0	99	0.0004	1322	0	76
DBCM	0.0130	552	62	13	0.0130	466	37	8.4	0.0002	350	0	87	0.0002	781	0	80	0.0004	1333	0	49
Bromoform	0.0270	552	362	1.4	0.0270	466	77	8.2	0.0004	330	0	76	0.0004	774	0	84	0.0011	1310	1	42
Others																				
1,4-DCB	0.0730	552	35	91	0.0730	466	68	80	0.0412	304	17	83	0.0849	807	5	51	0.0849	1322	2	54
PERC	0.0300	552	355	30	0.0300	466	97	55	0.0144	286	3	76	0.0339	978	1	33	0.0339	1317	0	17

MDL, method detection limit (μg L^{-1}); n, sample size (all measurements, including "extreme or illogical" values); Extre, numbers of "extreme or illogical" values.; DF, detection frequency (%).

Supplemental Table S 3. Results of linear mixed-effect models for outdoor VOCs in RIOPA using multiply imputed datasets (n = 2,775).

Benzene						1,4-DCB						PERC					
Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change
Intercept		0.51	0.34	0.138	30.4	Intercept		-0.55	0.21	0.011	8.8	Intercept		-2.34	0.18	<.0001	3.4
Visit	1	-0.08	0.05	0.141	126.9	Visit	1	0.03	0.07	0.693	11.7	Visit	1	-0.07	0.06	0.255	127.0
	2	Reference					2	Reference					2	Reference			
City	CA	-0.56	0.10	<.0001	-0.8	City	CA	0.78	0.11	<.0001	-2.4	City	CA	1.45	0.10	<.0001	3.3
	NJ	-0.62	0.09	<.0001	-18.7		NJ	0.59	0.13	<.0001	-6.5		NJ	1.36	0.09	<.0001	13.2
	TX	Reference					TX	Reference					TX	Reference			
Inverse wind speed	$knot^{-1}$	4.18	0.46	<.0001	0.1	Number of floors		0.10	0.04	0.009	3.2	Inverse wind speed	$knot^{-1}$	4.61	0.51	<.0001	-0.4
Ambient relative humidity	%	-0.01	0.00	<.0001	-17.0	Outdoor temperature	Q1	-0.44	0.12	0.000	20.0	No pets	No	-0.15	0.08	0.062	-29.2
Outdoor temperature	Q1	0.40	0.08	<.0001	2.7		Q2	-0.27	0.11	0.017	-13.3		Yes	Reference			
	Q2	0.31	0.08	<.0001	-5.0		Q3	-0.03	0.11	0.800	-195.8	Vacuuming	No	0.18	0.06	0.004	13.2
	Q3	-0.01	0.08	0.926	-22.3		Q4	Reference					Yes	Reference			
	Q4	Reference				Furniture refinisher in neighborhood	No	-0.68	0.20	0.001	-11.5	Dry cleaners in neighborhood	No	-0.12	0.07	0.076	-25.8
Near diesel vehicles	No	-0.20	0.06	0.000	-1.0		Yes	Reference					Yes	Reference			
	Yes	Reference				Airconditioning	No	0.31	0.09	0.001	-13.2	Number of carpeted rooms		-0.05	0.02	0.010	8.3
Gardening	No	0.16	0.08	0.047	-5.1		Yes	Reference					Q1	0.11	0.09	0.242	-28.6
	Yes	Reference				Using deordorizers or fresheners	No	0.16	0.09	0.087	9.4	Outdoor temperature	Q2	0.36	0.09	<.0001	8.9
Crawl space	No	-0.15	0.08	0.078	-16.5		Yes	Reference					Q3	0.03	0.09	0.700	2099.0
	Yes	Reference											Q4	Reference			
Tobacco products smoked in home	No	0.39	0.21	0.062	-40.8												
	Yes	Reference															
Ethnicity	White	-0.19	0.10	0.067	-4.9												
	Mexican	0.06	0.12	0.635	-41.8												
	Hispanic	-0.16	0.11	0.156	659.7												
	Other	Reference															

Supplemental Table S 4. Results of linear mixed-effect models for indoor VOCs in RIOPA using multiply imputed datasets (n = 2,770).

Benzene						PERC						α-Pinene					
Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change
Intercept		2.67	0.36	<.0001	4.0	Intercept		-1.71	0.22	<.0001	-13.9	Intercept		2.50	0.20	<.0001	0.9
Visit	1	-0.17	0.07	0.008	-19.3	Visit	1	-0.02	0.07	0.752	-141.5	Visit	1	0.09	0.06	0.143	-21.9
	2	Reference					2	Reference					2	Reference			
City	CA	-0.56	0.12	<.0001	7.4	City	CA	1.05	0.15	<.0001	7.6	City	CA	-0.47	0.12	<.0001	1.0
	NJ	-0.77	0.10	<.0001	-5.1		NJ	1.11	0.12	<.0001	-7.4		NJ	-0.60	0.13	<.0001	-2.5
	TX	Reference					TX	Reference					TX	Reference			
Number of rooms		-0.08	0.02	0.000	-12.5	Inverse wind speed	$knot^{-1}$	3.09	0.63	<.0001	-22.7	Number of rooms		-0.07	0.03	0.011	-5.7
Unemployed	No	0.10	0.09	0.298	-41.7	Visited dry cleaners during past week	No	-0.33	0.13	0.012	-1.1	Other members of the household took showers	No	-0.54	0.11	<.0001	-2.5
	Yes	Reference					Yes	Reference					Yes	Reference			
Education	Less than HS	0.27	0.12	0.024	-22.7	Sweeping indoors	No	0.15	0.09	0.088	-6.2	Using central air conditioning	No	-0.67	0.10	<.0001	7.6
	High school	0.04	0.10	0.719	3.8		Yes	Reference					Yes	Reference			
	College or above	Reference				Cooking inside or outside	No	0.18	0.08	0.027	-13.6	Logtransformed AER	hr^{-1}	-0.44	0.05	<.0001	-4.4
Professional cleaning	No	0.17	0.10	0.088	-11.7		Yes	Reference				Spending awake time at	1st floor	-0.39	0.11	0.001	-1.1
	Yes	Reference				Vacuuming	No	0.21	0.08	0.013	-21.1		Others	Reference			
Indoor temperature	°C	-0.04	0.01	0.000	12.0		Yes	Reference									
Attached garage	No	-0.19	0.09	0.029	-16.9	Vinyl, asbestos or other siding	No	0.27	0.11	0.015	-30.2						
	Yes	Reference					Yes	Reference									
						Professional cleaning	No	-0.14	0.13	0.292	-50.0						
							Yes	Reference									
						Logtransformed AER	hr^{-1}	-0.25	0.05	<.0001	-16.6						
						Unemployed	No	0.20	0.11	0.057	-15.7						
							Yes	Reference									

Supplemental Table S 5. Results of linear mixed-effect models for personal VOCs in RIOPA using multiply imputed datasets (n = 2,720).

Benzene						Styrene						d-Limonene					
Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change	Variable	Group	β	SE	p-value	% change
Intercept		2.51	0.38	<.0001	13.7	Intercept		1.00	0.34	0.003	-8.2	Intercept		3.34	0.36	<.0001	-7.8
Visit	1	-0.07	0.06	0.282	119.3	Visit	1	0.07	0.08	0.331	-2.1	Visit	1	-0.01	0.11	0.898	-115.0
	2	Reference					2	Reference					2	Reference			
City	CA	-0.80	0.11	<.0001	-3.8	City	CA	-0.21	0.11	0.060	-7.2	City	CA	-0.77	0.18	<.0001	-5.9
	NJ	-0.37	0.12	0.002	0.8		NJ	-0.10	0.10	0.320	-8.1		NJ	-0.96	0.17	<.0001	-14.3
	TX	Reference					TX	Reference					TX	Reference			
Inverse wind speed	$knot^{-1}$	3.60	0.53	<.0001	-14.3	Number of rooms		-0.09	0.02	0.000	-7.0	Number of rooms		-0.09	0.04	0.011	-29.4
Number of rooms		-0.10	0.02	<.0001	-0.9	Time spent indoors at home	min	0.00	0.00	0.004	-5.6	Other members of the household took showers	No	-0.74	0.17	<.0001	-6.8
Number of floors		-0.13	0.03	0.000	-16.2	Open doors or windows	No	0.22	0.09	0.014	13.2		Yes	Reference			
Heating fuel	Electricity	0.10	0.17	0.558	-47.7		Yes	Reference				Logtransformed AER	hr^{-1}	-0.34	0.08	<.0001	3.8
	Gas	0.31	0.15	0.038	-25.7	Spent at least 15 minutes in an enclosed garage with a parked car	No	-0.41	0.25	0.100	-1.5	Renovation to the house in the past year	No	-0.32	0.15	0.043	-30.1
	Oil and wood	Reference					Yes	Reference					Yes	Reference			
Indoor temperature	°C	-0.05	0.01	<.0001	11.8							Unemployed	No	-0.40	0.14	0.005	15.2
Education	Less than HS	0.13	0.12	0.288	-12.8								Yes	Reference			
	High school	-0.04	0.11	0.696	-48.8							Using other heaters (no central heating system)	No	0.53	0.25	0.035	-3.4
	> College	Reference											Yes	Reference			
Attached garage	No	-0.18	0.09	0.050	-4.3												
	Yes	0.00	.	.													
Pumping gas	No	-0.17	0.08	0.044	4.3												
	Yes	Reference															

Supplemental Table S 6. Statistics of VOC concentrations (μg L^{-1}) in blood measured for each NHANES cohort.

VOCs	NHANES 1988-1991				NHANES 1991-1994				NHANES 1999/2000				NHANES 2001/2002				NHANES 2003/2004			
	n	Mean	SE	50^{th}	n	Mean	SE	50^{th}	n	Mean	SE	50^{th}	n	Mean	SE	50^{th}	n	Mean	SE	50^{th}
Aromatics																				
Benzene	418	0.147	0.003	0.065	378	0.117	0.010	0.061	300	0.184	0.015	0.103	837	0.082	0.021	0.027	1345	0.069	0.004	0.028
Toluene	190	0.510	0.012	0.291	385	0.628	0.089	0.275	298	0.420	0.023	0.234	953	0.291	0.054	0.152	1336	0.216	0.018	0.091
Ethylbenzene	200	0.111	0.002	0.054	406	0.131	0.016	0.063	261	0.074	0.007	0.042	879	0.046	0.008	0.029	1299	0.044	0.001	0.031
m,p-Xylene	552	0.195	0.103	0.023	466	0.302	0.011	0.185	294	0.256	0.013	0.174	962	0.225	0.053	0.150	1346	0.168	0.008	0.130
o-Xylene	199	0.122	0.001	0.099	429	0.165	0.012	0.102	308	0.070	0.008	0.038	981	0.057	0.008	0.035	1365	0.045	0.002	0.035
BTEX	552	0.525	0.150	0.161	466	1.193	0.112	0.680	320	0.922	0.054	0.563	1015	0.652	0.117	0.390	1368	0.535	0.029	0.320
Styrene	200	0.158	0.001	0.042	424	0.070	0.006	0.040	283	0.067	0.004	0.042	948	0.092	0.009	0.024	1245	0.043	0.003	0.021
THMs																				
Chloroform	443	0.045	0.004	0.024	433	0.040	0.005	0.023	253	0.058	0.005	0.033	741	0.026	0.004	0.017	1222	0.020	0.005	0.010
BDCM	511	0.008	0.000	0.006	426	0.008	0.000	0.006	354	0.004	0.000	0.002	785	0.004	0.001	0.002	1322	0.003	0.000	0.002
DBCM	490	0.011	0.001	0.009	429	0.010	0.000	0.009	350	0.003	0.000	0.001	781	0.002	0.000	0.001	1333	0.002	0.000	0.000
Bromoform	190	0.021	0.000	0.019	389	0.021	0.000	0.019	330	0.002	0.000	0.001	774	0.004	0.001	0.001	1309	0.004	0.001	0.001
∑THM	551	0.059	0.007	0.049	465	0.072	0.006	0.050	356	0.047	0.005	0.028	820	0.033	0.003	0.025	1337	0.026	0.004	0.014
Others																				
1,4-DCB	517	1.145	0.098	0.294	398	1.071	0.192	0.374	287	0.875	0.230	0.219	802	0.935	0.305	0.087	1320	0.827	0.142	0.140
PERC	197	0.142	0.004	0.075	369	0.255	0.025	0.055	283	0.110	0.014	0.043	977	0.070	0.004	0.034	1317	0.081	0.011	0.034

Sample size n including valid measurements; values below MDL measurements were replaced by 1/2 MDL.
SE, standard error.

Supplemental Table S 7. Predicted excess cancer risk for RIOPA adult participants (n = 239).

VOCs	Unit risk $(\mu g\ m^{-3})^{-1}$	Predicted excess cancer cases per million population						
		Mean	SD	Min	25th	75th	98th	Max
Benzene	7.8×10^{-6}	28.4	25.9	4.3#	13.5	32.7	134.2	172.6
Ethylbenzene	2.5×10^{-6}	7.1	9.9	0.9#	3.0	7.6	43.2	82.9
MTBE	2.6×10^{-7}	3.5	4.6	0.1#	1.2	4.1	17.5	37.2
Styrene	2.0×10^{-6}	3.2	6.9	0.3#	0.8#	2.6	23.9	59.9
1,4-DCB	1.1×10^{-5}	626.5	2223.0	2.4#	10.0#	126.0	9518.1	19167.0
TCE	2.0×10^{-6}	1.4	4.1	0.2#	0.2#	0.93	16.1	40.9
PERC	5.9×10^{-6}	12.9	25.9	0.7#	2.5#	11.8	97.5	242.3
Chloroform	2.3×10^{-5}	47.0	62.2	3.2#	14.5	52.6	248.8	537.6
CTC	1.5×10^{-5}	9.8	2.9	2.0#	8.2	10.7	17.1	27.8
Hematopoietic mixture	NA	680.2	2239.7	12.78	44.89	180.22	9695.8	19195.8
Liver and kidney toxicant mixture	NA	714.8	2247.4	20.80	61.25	265.03	9723.1	19222.9
Total VOC	NA	745.8	2253.9	34.1	83.9	293.3	9780.5	19250.0

NA, not available; SD, standard deviation; min, minimum; max, maximum.
#, concentrations were based on MDLs.
Hematopoietic mixture includes benzene, MTBE, 1,4-DCB, TCE and PERC; liver and kidney toxicant mixture includes ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform and CTC.

Supplemental Table S 8. Linear quantile regressions of log-transformed blood VOC concentrations for NHANES 1988-1994 and 2001-2004 (without 1999/2000).*

VOCs	Quantile 0.5			Quantile 0.75			Quantile 0.95		
	β	SE	% change	β	SE	% change	β	SE	% change
Aromatics									
Benzene	**-0.043**	0.003	-20.1#	**-0.038**	0.004	-51.9#	-0.023	0.014	NA
Toluene	**-0.071**	0.005	-27.7#	**-0.089**	0.012	-38.1#	**-0.089**	0.012	-12.9
Ethylbenzene	**-0.054**	0.006	-9.2	**-0.055**	0.006	-15.7	**-0.095**	0.027	-19.4
m,p-Xylene	**-0.024**	0.006	-28.4	**-0.055**	0.008	-4.5	**-0.136**	0.049	16.0
o-Xylene	**-0.082**	0.000	18.5#	**-0.098**	0.007	0.7	**-0.129**	0.035	5.4
BTEX	**-0.043**	0.005	-35.3#	**-0.059**	0.008	-26.6	**-0.083**	0.029	17.5
Styrene	**-0.029**	0.004	-18.7	**-0.025**	0.006	-35.0	**-0.096**	0.039	56.7
THMs									
Chloroform	**-0.059**	0.004	-9.3	**-0.054**	0.006	-15.2	**-0.059**	0.015	-42.4
BDCM	**-0.102**	0.007	5.4	**-0.040**	0.001	-6.5	**-0.034**	0.010	1.2
DBCM	**-0.202**	0.018	0.0	**-0.142**	0.004	-4.6	**-0.077**	0.007	0.0
Bromoform	**-0.241**	0.000	0.0	**-0.196**	0.001	-2.4#	**-0.161**	0.028	25.9
∑THM	**-0.112**	0.006	-2.4	**-0.092**	0.005	-8.8	**-0.052**	0.020	-55.1
Others									
1,4-DCB	**-0.061**	0.001	-2.6	**-0.040**	0.009	-11.0	-0.031	0.022	NA
PERC	NA	NA	NA	**-0.149**	0.003	-10.3#	**-0.144**	0.038	-18.5

*, excludes 1988-1991 data for toluene, ethylbenzene, m,p-xylene, o-xylene, BTEX, styrene, bromoform, ∑THM and PERC.
Aromatic VOCs were adjusted for solvent-related occupations and serum cotinine levels; THMs and other VOCs were adjusted for solvent-related occupations only.
SE=standard error; NA=not available.
% change=(((βwithout1999/2000-βwith1999/2000)/βwith1999/2000)100%), which was calculated only when both βwith1999/2000 and βwithout1999/2000 were significant.
Bold type means statistically significant ($p<0.05$) in QR models; # means differences between βwith1999/2000 and βwithout1999/2000 were statistically significant examined by approximately Wald tests.

Supplemental Table S 9. Piecewise quantile regressions using knot at 1999/2000 for log-transformed blood VOC concentrations in the 1988 to 2004 period.*

VOCs	0.5 Quantile		0.75 Quantile		0.95 Quantile	
	Slope1 (SE)	Slope2 (SE)	Slope1 (SE)	Slope2 (SE)	Slope1 (SE)	Slope2 (SE)
Aromatics						
Benzene	**0.061 (0.009)**	**-0.244 (0.011)**	**0.082 (0.008)**	**-0.260 (0.010)**	**0.067 (0.022)**	**-0.204 (0.030)**
Toluene	0.018 (0.011)	**-0.207 (0.012)**	**0.052 (0.018)**	**-0.238 (0.016)**	-0.040 (0.042)	**-0.176 (0.040)**
Ethylbenzene	**-0.059 (0.013)**	**-0.060 (0.014)**	**-0.032 (0.012)**	**-0.090 (0.015)**	-0.051 (0.057)	**-0.135 (0.029)**
m,p-Xylene	0.008 (0.012)	**-0.069 (0.012)**	0.016 (0.011)	**-0.116 (0.012)**	**0.107 (0.025)**	**-0.263 (0.030)**
o-Xylene	**-0.148 (0.005)**	**-0.006 (0.002)**	**-0.067 (0.013)**	**-0.127 (0.016)**	-0.039 (0.053)	**-0.163 (0.032)**
BTEX	0.009 (0.011)	**-0.126 (0.009)**	0.020 (0.014)	**-0.154 (0.013)**	0.010 (0.037)	**-0.124 (0.030)**
Styrene	-0.001 (0.006)	**-0.067 (0.006)**	**0.051 (0.012)**	**-0.134 (0.014)**	**0.113 (0.036)**	**-0.168 (0.030)**
THMs						
Chloroform	**0.062 (0.013)**	**-0.294 (0.017)**	**0.084 (0.011)**	**-0.301 (0.018)**	**0.039 (0.018)**	**-0.273 (0.024)**
BDCM	**-0.084 (0.009)**	**-0.124 (0.021)**	-0.011 (0.012)	**-0.094 (0.010)**	-0.013 (0.021)	**0.076 (0.033)**
DBCM	**-0.168 (0.013)**	**-0.268 (0.021)**	**-0.116 (0.015)**	**-0.205 (0.027)**	**-0.077 (0.021)**	**-0.076 (0.036)**
Bromoform	**-0.403 (0.000)**	**-0.014 (0.000)**	**-0.306 (0.005)**	-0.032 (0.037)	**-0.200 (0.049)**	-0.028 (0.060)
∑THM	**-0.054 (0.013)**	**-0.184 (0.014)**	-0.003 (0.011)	**-0.215 (0.013)**	0.044 (0.031)	**-0.195 (0.027)**
Others						
1,4-DCB	**-0.067 (0.007)**	**-0.056 (0.012)**	-0.024 (0.020)	**-0.072 (0.027)**	0.012 (0.042)	-0.123 (0.075)
PERC	**-0.121 (0.000)**	0.000 (0.000)	**-0.110 (0.025)**	**-0.191 (0.004)**	-0.109 (0.077)	**-0.207 (0.057)**

*, excludes 1988-1991 data for toluene, ethylbenzene, m,p-xylene, o-xylene, BTEX, styrene, bromoform, ∑THM and PERC.
Aromatic VOCs were adjusted for solvent-related occupations and serum cotinine levels; THMs and other VOCs were adjusted for solvent-related occupations.
Slope1=slope of regression line connecting 1988-1991 and 1999/2000; slope2=slope of regression line connecting 1999/2000 and 2003/2004.
SE=standard error; NA=not available.
Bold type means statistically significant ($p<0.05$).

Supplemental Table S 10. Effect sizes* of linear mixed-effect models for personal exposure to gasoline-related VOCs in RIOPA.

Variable	Group/unit	Benzene		Toluene		Ethylbenzene		m,p-Xylene		o-Xylene		MTBE		Styrene	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Intercept		**9.13**	**2.21**	**42.16**	**2.05**	**4.10**	**2.27**	**9.32**	**2.06**	**2.18**	**1.78**	**6.18**	**1.85**	**2.98**	**1.93**
Visit	1	1.031	1.136	1.13	1.19	-1.15	1.17	-1.09	1.18	-1.07	1.16	1.06	1.22	1.08	1.16
	2	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
City	Los Angeles	**-2.29**	**1.26**	1.09	1.24	**-1.45**	**1.31**	**-1.34**	**1.32**	-1.06	1.30	**-1.41**	**1.38**	**-1.26**	**1.25**
	Elizabeth	**-1.44**	**1.32**	1.07	1.29	-1.17	1.43	-1.29	1.45	-1.19	1.39	1.07	1.48	-1.12	1.23
	Houston	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
Attached garage	No	**-1.21**	**1.19**	**-2.06**	**1.63**	**-1.44**	**1.26**	**-1.43**	**1.27**	**-1.42**	**1.23**	**-1.43**	**1.27**	-1.51	1.63
Cooking	No			**1.24**	**1.19**	**1.19**	**1.18**	1.17	1.19	**1.22**	**1.16**				
Education	Less than HS	1.16	1.27												
	High school	-1.09	1.22												
	> College	Reference													
Ethnicity	White			-1.14	1.35			-1.25	1.37	-1.23	1.32				
	Mexican			1.21	1.44			1.07	1.46	1.12	1.40				
	Hispanic			1.35	1.45			1.31	1.48	**1.42**	**1.41**				
	Other					Reference		Reference		Reference					
Heating fuel	Electricity	1.22	1.42												
	Gas	**1.52**	**1.37**												
	Oil and wood	Reference													
Indoor temperature	°C	**-1.17**	**1.08**												
Inverse wind speed	$knot^{-1}$	**1.52**	**1.11**			**1.37**	**1.14**	**1.33**	**1.15**	**1.29**	**1.13**	**1.80**	**1.18**		
Log-transformed AER	hr^{-1}			**-1.39**	**1.12**	**-1.20**	**1.13**	**-1.26**	**1.13**	**-1.17**	**1.12**	**-1.10**	**1.15**		
Number of floors		**-1.35**	**1.16**									**-1.48**	**1.26**		
Number of rooms		**-1.21**	**1.10**											**-1.21**	**1.10**
Open doors or windows	No									**1.25**	**1.20**			**1.22**	**1.20**
Pumping gas	No	**-1.18**	**1.18**			**-1.27**	**1.25**	**-1.24**	**1.25**	**-1.32**	**1.22**	**-1.40**	**1.30**		
Renovation in the past year	No			**-1.35**	**1.22**										
Time spent in home	min			**-1.18**	**1.16**	-1.15	1.16							**-1.22**	**1.13**
Unemployed	No											1.26	1.27		
Using air cleaning devices	No					-1.31	1.42	**-1.52**	**1.43**	**-1.46**	**1.37**	-1.42	1.49		
Using nail polish remover	No			-1.34	1.39	**-1.48**	**1.38**	**-1.38**	**1.40**						
Wore powder, spray or perfume	No											**1.50**	**1.26**		

*, for continuous variables, the effect size ($\mu g\ m^{-3}$) is equal to the change in exposure for one inter-quartile range of the determinant.
AER, air exchange rate; HS, high school.
For dichotomous variables, the reference group is "Yes".
p-value < 0.05 shown in bold type.

Supplemental Table S 11. Effect sizes* of linear mixed-effect models for personal exposure to odorant-related VOCs in RIOPA.

Variable	Group/unit	1,4-DCB		Chloroform		d-Limonene		α-Pinene		β-Pinene	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Intercept		**33.23**	**4.60**	**3.83**	**2.53**	**37.39**	**2.14**	**11.27**	**1.62**	**4.80**	**2.36**
Visit	1	**1.40**	**1.33**	1.17	1.19	1.10	1.34	**1.19**	**1.16**	1.08	1.21
	2	Reference		Reference		Reference		Reference		Reference	
City	Los Angeles	**-3.00**	**1.79**	**-1.56**	**1.36**	**-2.27**	**1.44**	**-2.04**	**1.28**	**-3.18**	**1.34**
	Elizabeth	**-2.25**	**1.82**	-1.06	1.40	**-3.07**	**1.54**	**-1.81**	**1.31**	**-2.88**	**1.39**
	Houston	Reference		Reference		Reference		Reference		Reference	
Air conditioning	No	**1.71**	**1.56**					**-1.67**	**1.23**	-1.22	1.28
Ambient relative humidity	%			**-1.14**	**1.12**					**-1.14**	**1.12**
Furniture refinisher in neighborhood	No	**-3.66**	**2.65**								
Waxing or polishing furniture	No	**-2.24**	**1.90**								
Keeping dogs or cats	No							1.17	1.22	**1.34**	**1.24**
Log-transformed AER	hr^{-1}			**-1.56**	**1.15**	**-1.43**	**1.19**	**-1.54**	**1.12**	**-1.40**	**1.15**
Not using fresheners or candles	No									1.37	1.42
Number of rooms		**-1.32**	**1.30**	**-1.26**	**1.18**	**-1.29**	**1.18**	**-1.21**	**1.13**		
Open doors or windows	No	**1.52**	**1.47**							1.24	1.27
Other family members took showers	No			**-1.47**	**1.34**	**-2.22**	**1.43**	**-1.51**	**1.27**	**-1.42**	**1.32**
Outdoor swimming pool or hot tub	No							**-1.37**	**1.28**		
Using heating at	< 64 °F	**2.14**	**1.68**								
	64 to 70 °F	-1.03	1.59								
	> 70 °F	Reference									
Ownership of the house	No			**1.34**	**1.33**						
Pets indoors	No			**1.37**	**1.26**						
Renovation in the past year	No					**-1.57**	**1.34**				
Restaurants or bakery in neighborhood	No	**-1.87**	**1.70**								
Unemployed	No					**-1.42**	**1.36**				
Using a clothes washer	No	**1.70**	**1.46**								
Using dishwashers	No			**-1.29**	**1.30**						
Using other heaters (non-CHS)	No					**1.73**	**1.68**				

*, for continuous variables, the effect size (μg m^{-3}) is equal to the change in exposure for one inter-quartile range of the determinant.
AER, air exchange rate; HS, high school.
For dichotomous variables, the reference group is "Yes".
p-value < 0.05 shown in bold type.

Supplemental Table S 12. Effect sizes* of linear mixed-effect models for personal exposure to dry-cleaning and industrial-related VOCs in RIOPA.

Variable	Group/unit	TCE		PERC		CTC	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Intercept		**-2.21**	**2.29**	-1.62	2.64	**-1.89**	**1.58**
Visit	1	**1.20**	**1.15**	**1.21**	**1.21**	-1.01	1.07
	2	Reference		Reference		Reference	
City	Los Angeles	**1.94**	**1.33**	**1.78**	**1.42**	**-1.19**	**1.15**
	Elizabeth	**3.42**	**1.33**	**1.71**	**1.60**	-1.12	1.16
	Houston	Reference		Reference		Reference	
Ambient relative humidity	%			**-1.13**	**1.13**		
Ethnicity	White			-1.13	1.45		
	Mexican			**-1.62**	**1.57**		
	Hispanic			1.06	1.60		
	Other			Reference			
Having a fireplace	No					**-1.14**	**1.14**
Indoor temperature	°C	**-1.10**	**1.10**			1.04	1.04
Inverse wind speed	$knot^{-1}$			**1.63**	**1.18**		
Log-transformed AER	hr^{-1}			**-1.24**	**1.15**		
Not using fresheners or candles	No					**-1.22**	**1.16**
Restaurants or bakery in neighborhood	No	**1.30**	**1.30**				
Source of household water	Public	**-1.78**	**1.69**			**1.65**	**1.32**
Sweeping indoors	No			1.21	1.26		
Time spent at closed cars	min	**1.25**	**1.12**				
Unemployed	No			**1.52**	**1.28**		
Using air cleaning devices	No					**-1.21**	**1.18**
Vinyl, asbestos or other siding	No	**-1.28**	**1.29**				
Visited dry cleaners during past week	No			**-1.88**	**1.34**		

*, for continuous variables, the effect size ($\mu g\ m^{-3}$) is equal to the change in exposure for one inter-quartile range of the determinant.
AER, air exchange rate; HS, high school.
For dichotomous variables, the reference group is "Yes".
p-value < 0.05 shown in bold type.

CHAPTER 4

Conclusions

This dissertation draws on the outdoor, indoor, personal and biological VOC measurements from two large datasets, RIOPA and NHANES, and utilizes several novel and powerful statistical modeling and analysis techniques. It identifies and characterizes exposure distributions, risks, trends, mixtures, dependencies of the components in mixtures, and exposure determinants. The conclusions are presented in this chapter. Section 4.1 summarizes the main findings for each objective (see Section 1.3). Section 4.2 addresses the relevance of the findings to public health and environmental concerns. Section 4.3 suggests possible applications of the advanced statistical methods used in this research, and identifies unsolved scientific issues for further investigation.

4.1 Main Findings

4.1.1 Extreme Value Analyses

The results of the extreme value analyses (Section 3.3) showed that the highest exposures in RIOPA, which can be the most significant in terms of health risks, closely fitted generalized extreme value (GEV) distributions and, in many cases, Gumbel distributions, a reduced form of the GEV distribution. In contrast, lognormal distributions, the usual "default" distributional assumption, underestimated concentrations and risks from extrema. 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 predict the numbers of individuals that exceed risk-based exposure guidelines or other criteria are needed. GEV distributions will be useful in impact and policy analyses to describe concentrations, exposures and risks.

4.1.2 Mixture of Normal Distributions

Although GEV distributions can represent tail behavior of exposure and risk distributions,

they do not fit the full distribution of most environmental data, which can have multiple modes, heavy tails, left-censoring, and other features. Compared to parametric distributions, the finite mixture of normals and Dirichlet process mixture (DPM) of normals were shown (Section 3.4) to have superior performance in fitting VOC exposure data with heavy tails or with a large fraction of data below the method detection limits (MDLs). The optimal number of distributions (k) needed for the finite mixture of normals models ranged from 2 to 4, depending on VOCs. Distributions from the DPMs provided slightly better fits than the finite mixture of normals. This 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 under left censoring (due to the MDL).

4.1.3 Trend Analyses

In Section 3.5, VOC exposure trends from 1988 to 2004 were examine using concentrations measurements in blood drawn from five cohorts of NHANES, a large and nationally representative sample of U.S. adults. There is no question that VOC exposures decreased over this period, however, the rate of decrease depends on the both the VOC and the quantile. Using quantile regression (QR) models, three patterns were discerned: exposures of benzene, toluene, BTEX and, with less confidence, ΣTHMs and chloroform, had similar decreases at all quantiles (pattern 1); ethylbenzene, m,p-xylene, o-xylene, styrene and PERC levels decreased fastest at upper quantiles (pattern 2); and 1,4-DCB declined faster at central quantiles (pattern 3). Because the sample included participants with a wide range of occupations and exposures, upper quantile exposures may reflect occupational exposure, while lower quantiles arise from general environmental sources. There is less certainty regarding the nature of the exposure trends. Linear models yielded reductions of 2.5 to 6.4% per year for most VOCs, a robust result that is consistent with ambient trends, described below. Shorter term trends, evaluated using piecewise models and other analyses, suggest that several VOCs had smaller changes through the 1990's, followed by swifter reductions in subsequent years; however, these trends may be driven by previously unreported anomalies in the NHANES data that affected the 1988 through 2000 cohorts.

VOC emissions and ambient concentrations were compared to the biomonitoring data. For most VOCs, reported emissions decreased more slowly (e.g., 4-6% per year for toluene and xylene from 1999 to 2004) than median exposures. However, for most VOCs, long term trends of ambient concentrations decreased more rapidly than the NHANES exposure data. Exposure, emission and concentration trends may diverge, especially for VOCs with strong indoor sources, e.g., chloroform and 1,4-DCB. These differences suggest the importance of indoor emission sources, smoking, occupation, personal activities and other factors on exposure, in addition to emissions and ambient concentrations.

Internal checks on the validity of the NHANES measurements were made by comparing blood and personal sampling measurements collected in the 1999/2000 cohort, and by comparing results across cohorts. The low to moderate correlation found can be explained by NHANE's experimental design, the rapid clearance of most VOCs from blood, and other factors. It should be noted that data were insufficient to estimate trends for BDCM, DBCM and bromoform, and also that portions of the 1988-1991 through 1999/2000 VOC data appear unreliable. Still, the NHANES measurements are unique and valuable in providing a 15 year history of population exposure to VOCs in the U.S.

4.1.4 Identification of Mixtures

Many VOCs have similar emission sources and/or toxicological effects, highlighting the need to understand and evaluate exposures to mixtures. VOC mixtures in the RIOPA dataset were identified using positive matrix factorization (PMF) analyses and the toxicological mode of action (Sections 2.2.6.2 and 3.6). 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 air exchange rates. The influence of environmental factors and personal activities was also shown for certain mixtures, e.g., mixture A1 was associated with attached garages and self-service pumping gas. Three additional mixtures based on cancer endpoints were identified, which respectively can cause liver and renal tumors (mixtures B1 and A3/B3), and hematopoietic cancers (mixture B2).

4.1.5 Dependencies of Components in Mixtures

Dependencies between mixture components were described using copulas (Section 3.7), 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, while two other mixtures best fitted Gumbel copulas, which better 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 poorly fit lognormal distributions, and which may represent the greatest risk.

4.1.6 Exposure Determinants

LMMs were used to identify determinants of VOC exposures in RIOPA (Section 3.9). The determinants 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). Most of these factors were associated with indoor concentrations in the participant's home, which contributed a large share exposure (average exposure fractions ranged from 63% for MTBE to 75% for CTC). Gasoline-, odorant and cleaning-, and dry-cleaning and industry-related VOCs were associated with a number of individual and environmental determinants, consistent with previous studies, e.g., gasoline-related VOCs were higher in homes with attached garages, and dry cleaning-related VOCs were higher in participants who visited dry cleaners. Several new determinants were identified, including effects of city, other family member showering, and residence size. Outdoor VOC concentrations provided small contributions to VOC exposure (exposure fractions averaged from 0.032 to 0.006). To extend and generalize results, further investigation using a more representative population and a wider suite of VOCs is suggested.

4.2 Implications of Findings

This dissertation 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 and NHANES VOC datasets. These methods performed well, and they deserve more widespread consideration and application.

First, the highest exposure events do not fit “default” distributional assumptions, i.e., 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 the need to more accurately characterize and model these high concentrations and exposures, potentially using the extreme value theory, and that the use of this enhanced information and methods for estimating population risks and establishing exposure and risk guidelines.

Second, single (parametric) distributions may not accurately fit exposure data, which contains features such as multiple modes, heavy tails, and left censoring. The suggested mixture models, finite mixture of normals and DPM of normals, provided much better fits to the RIOPA VOC dataset than 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 and/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.

Third, trends of VOC exposures were evaluated using QR models and 1988 to 2004 NHANES data. This analysis reveals changes in blood VOC levels in the U.S. population (20 to 59 year old) over past decades. The trends were examined at various percentiles (one of the greatest advantages of using QR models), and showed different patterns, which may reflect changes in exposure sources. Additionally, exposure trends were compared to trends of emissions and ambient VOCs. The results reflect declining trends in emissions and ambient VOC levels, but also suggest the importance of indoor sources and personal activities on VOC exposures.

Fourth, copulas were used to estimate dependency structures in mixtures of VOCs. The RIOPA dataset showed complex dependencies, e.g., 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 they decouple the dependency structure from the marginal distribution. These are essential considerations for cumulative exposure and cumulative risk assessment, and copulas provide a powerful tool in this application, especially for high concentration mixtures that may pose the greatest risks.

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

4.3 Recommendations for Further Study

This dissertation used data drawn from RIOPA and NHANES, much of which was collected over a decade ago. Updated data are needed to explore and understand current exposure situations. For example, the most recent blood VOC data in NHANES was from the 2005/2006 cohort (latest release). Since this time, the survey has been expanded to include younger participants (from 12 year old). Further research could examine more recent trends of VOC exposures, and separate children and adult populations.

This dissertation has applied several advanced statistical methods, but these methods rarely have been applied in other environmental studies. Further applications of these methods are warranted. For example, considering the tail dependencies of VOC mixtures and the extreme value distributions of VOCs, future studies should apply extreme-value copulas (including Galambos and Husler-Reiss copulas, as well as Gumbel copulas), which combine the copula technique and the extreme value theory. Such approaches can predict the risk of exposure to extreme values of VOCs. In addition, dependency structures in VOC mixtures may change over time due to different emission sources or activity patterns, so the longitudinal NHANES data can be used with copulas to explore temporal joint distributions of VOC exposures. Also, copulas are recommended to estimate the dependency structures of other class of pollutants or across different types of pollutants.

Another possible application is the use of QR models for determining exposure factors. Determinants of VOC exposures may vary as a function of exposure levels, i.e., high- and low-exposed populations may be affected by different factors; further research could help explore exposure determinants at different percentiles using QR models.

Since people are typically exposed to mixtures, there will be a continuing need to estimate the determinants of such exposures. In this case, the associations between multiple correlated response variables (e.g., VOC mixtures), and covariates (e.g., potential determinants), can be estimated using copula regression models.

Finally, the general recommendation is that the statistical approaches used in these analyses are needed when investigating other pollutants like particulate matter, other settings such as other countries and other populations, especially sensitive populations, e.g., children and elders. This more comprehensive interpretation provides an improved foundation on which to base policy decisions.

REFERENCES

ACGIH. 2001. Documentation of the Biological Exposure Indices, 7th Ed. American Conference of Governmental Industrial Hygienists.

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

Adams, J C, Dills, R L, Morgan, M S, et al. 2005. A physiologically based toxicokinetic model of inhalation exposure to xylenes in Caucasian men. Regulatory Toxicology and Pharmacology 43(2): 203-2114.

Adgate JL, Eberly LE, Stroebel C, Pellizzari ED, Sexton K. 2004. Personal, indoor, and outdoor VOC exposures in a probability sample of children. Journal of Exposure Analysis and Environmental Epidemiology 14(S1): S4-S13.

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(4): 295-307.

Ang AH-S, Tang WH. 1975. Probability concepts in engineering planning and design: Decision, risk and reliabilit. New York: John Wiley and Sons.

Antoniak CE. 1974. Mixtures of Dirichlet Processes with applications to nonparametric problems. Annals of Statistics 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. Environmental science & technology 42(10): 3732-3738.

Ashley DL, Prah JD. 1997. Time dependence of blood concentrations during and after exposure to a mixture of volatile organic compounds. Archives of Environmental Health 52(1): 26-33.

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

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

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

ATSDR. 1997a. Toxicological profile for chloroform. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/toxprofiles/tp6.pdf.

ATSDR. 1997b. Toxicological profile for tetrachloroethylene. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/toxprofiles/tp18.pdf.

ATSDR. 2004. Guidance manual for the assessment of joint toxic action of chemical mixtures. Atlanta, GA: Agency for Toxic Substances and Disease Registry.

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

ATSDR. 2005b. Toxicological profile for naphthalene, 1-methylnaphthalene, and 2-methylnaphthalene. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available

from http://www.atsdr.cdc.gov/toxprofiles/tp67.pdf.
ATSDR. 2006a. Toxicological profile for dichlorobenzenes. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available from http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=704&tid=126.
ATSDR. 2006b. Toxicological profile for dichlorobenzenes. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/toxprofiles/tp10.pdf.
ATSDR. 2007. Toxicological profile for xylene. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available from http://www.atsdr.cdc.gov/toxprofiles/tp71.pdf.
ATSDR. 2007a. Toxicological profile for benzene. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/toxprofiles/tp3-p.pdf.
ATSDR. 2007b. Toxicological profile for ethylbenzene. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/ToxProfiles/tp110.pdf.
ATSDR. 2007c. Toxicological profile for styrene. Agency for Toxic Substances and Disease Registry, Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available from http://www.atsdr.cdc.gov/toxprofiles/tp53.pdf.
ATSDR. 2009. Toxicological profile for 1,3-butadiene. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available from http://www.atsdr.cdc.gov/ToxProfiles/tp28.pdf.
ATSDR. 2010a. Toxicological profile for ethylbenzene. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available from http://www.atsdr.cdc.gov/toxprofiles/tp110.pdf.
ATSDR. 2010b. Toxicological profile for styrene. Atlanta, GA: Agency for Toxic Substances and Disease Registry Available from http://www.atsdr.cdc.gov/PHS/PHS.asp?id=419&tid=74.
Azuma K, Uchiyama I, Ikeda K. 2007. The risk screening for indoor air pollution chemicals in Japan. Risk Analysis 27(6): 1623-1638.
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. Environmental Health Perspectives 105(8): 850-855.
Barnett V. 1975. Probability plotting methods and order statistics. Applied Statistics 24: 95-108.
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(4): 266-278.
Batterman S, Jia C, Godwin C, Hatzivasilis G. 2006. A dominant source of VOC exposure: Attached garages. Epidemiology 17(6): S350.
Batterman S, Jia C, Hatzivasilis G. 2007. Migration of volatile organic compounds from attached garages to residences: A major exposure source. Environmental Research 104(2): 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. Science of The Total Environment 409(6): 1058-1068.
Benaglia T, Chauveau D, Hunter DR, Young DS. 2009. mixtools: An R package for analyzing mixture models. Journal of Statistical Software 32(6).
Benowitz N. 1999. Biomarkers of environmental tobacco smoke exposure. Environ Health Perspect 107(Suppl 2): 349-355.
Biernacki C, Celeux C, Govaert G. 2000. Assessing a mixture model for clustering with the integrated completed likelihood. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(7): 719-725.
Blackwell D, MacQueen JB. 1973. Ferguson distributions via Polya Urn schemes. Annals of Statistics 1(2): 353-355.
Bliss C. 1939. The toxicity of poisons applied jointly. Annals of Applied Biology 26(3): 585-615.
Borgert CJ, Quill TF, McCarty LS, Mason AM. 2004. Can mode of action predict mixture toxicity for risk assessment? Toxicology and Applied Pharmacology 201(2): 85-96.

Boutonnet J-C, Thompson R, De Rooij C, Garny V, Lecloux A, Van Wijk D. 2004. 1,4-Dichlorobenzene marine risk assessment with special reference to the Osparcom Region: North Sea. Environmental Monitoring and Assessment 97(1): 103-117.
Brown SK. 2002. Volatile organic pollutants in new and established buildings in Melbourne, Australia. Indoor Air 12(1): 55-63.
Brugnone F, De Rosa E, Perbellini L, Bartolucci G. 1986. Toluene concentrations in the blood and alveolar air of workers during the workshift and the morning after. British Journal of Industrial Medicine 43: 56-61.
Brugnone F, Perbellini L, Maranelli G, Romeo L, Guglielmi G, Lombardini F. 1992. Reference values for blood benzene in the occupationally unexposed general population. International Archives of Occupational and Environmental Health 64(3): 179-184.
Burmaster DE, Wilson AM. 2000. Fitting second-order finite mixture models to data with many censored values using maximum likelihood estimation. Risk analysis : an official publication of the Society for Risk Analysis 20(2): 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. Journal of Environmental Monitoring 12(2): 524-535.
Cade BS, Noon BR. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1(8): 412-420.
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. Toxicology and Industrial Health 14(3): 429-454.
Cherubini U, Luciano E, Vecchiato W. 2004. Copula methods in finance. New York, NY: John Wiley and Sons.
Chin JY, Godwin C, Jia C, Robins T, Lewis T, Parker E, et al. 2012. Concentrations and risks of p-dichlorobenzene in indoor and outdoor air. Indoor Air doi: 10.1111/j.1600-0668.2012.00796.x.
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(1): 40-49.
Chu H, Kensler TW, Muñoz A. 2005. Assessing the effect of interventions in the context of mixture distributions with detection limits. Statistics in Medicine 24(13): 2053-2067.
Clobes AL, Ananth GP, Hood AL, Schroeder JA, Lee KA. 1992. Human activities as sources of volatile organic compounds in residential environments. Annals of the New York Academy of Sciences 641(1): 79-86.
D'Souza JC, Jia C, Mukherjee B, Batterman S. 2009. Ethnicity, housing and personal factors as determinants of VOC exposures. Atmospheric Environment 43(18): 2884-2892.
Delgado-Saborit JM, Aquilina NJ, Meddings C, Baker S, Harrison RM. 2009. Model development and validation of personal exposure to volatile organic compound concentrations. Environmental Health Perspectives 117(10).
Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological) 39(1): 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. Atmospheric Environment 35(27): 4531-4543.
Embrechts P, Klüppelberg C, Mikosch T. 1997. Modelling extremal events for insurance and finance: Springer.
Engeland K, Hisdal H, Frigessi A. 2004. Practical extreme value modelling of hydrological floods and droughts: A case study. Extremes 7(1): 5-30.
Escobar MD. 1994. Estimating normal means with a dirichlet process prior. Journal of the American Statistical Association 89(425): 268-277.
Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association 90(430): 577-588.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. The Annals of Statistics 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, U.S.: Academic Press, 287-302.
Feron V, Groten J, van Bladeren P. 1998. Exposure of humans to complex chemical mixtures: hazard identification and risk assessment. Archives of Toxicology Supplement 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. Mathematical Proceedings of the Cambridge Philosophical Society 24: 180-190.
Fraley C, Raftery AE. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. The Computer Journal 41(8): 578-588.
Frees E, Valdez E. 1998. Understanding relationships using copulas. North American Actuarial Journal 2(1): 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. Environmental Health Perspectives 114(4): 514-521.
Gumbel EJ. 1958. Statistics of extremes. Columbia University Press, New York.
Haas C. 1997. Importance of distributional form in characterizing inputs to Monte Carlo risk assessments. Risk Analysis 17(1): 107-113.
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: US 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(3): 533-541.
Hissink AM, Dunnewijk R, van Ommen B, van Bladeren PJ. 1997. Kinetics and metabolism of 1,4-dichlorobenzene in male Wistar rats: no evidence for quinone metabolites. Chemico-Biological Interactions 103(1): 17-33.
Hopke PK, Paatero P. 1994. Extreme-value estimation applied to aerosol size distributions and related environmental problems. Journal of Research of the National Institute of Standards and Technology 99(4): 361-367.
HSE. 2002. Guidelines for use of statistics for analysis of sample inspection of corrosion. UK: The Health and Safety Executive.
Huang Y-L, Batterman S. 2003. An extreme value analysis of pollutant concentrations in surface soils due to atmospheric deposition. Human and Ecological Risk Assessment 9(7): 1729-1746.
Hüsler J. 1983. "Frost data: A case study on extreme values of non-stationary sequences," in J. T. de Oliveira, (ed.). Boston, mA: Statistical Extremes and Applications, (D. Reidel Publishing Co.).
Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. Biometrika 76: 297-307.
IARC. 1982. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans: Some industrial chemicals and dyestuffs.: Lyon, France: International Agency for Research on Cancer.
IARC. 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.
IARC. 2011. Agents classified by the IARC monographs. Lyon, France: World Health Organization, International Agency for Research on Cancer Available from http://monographs.iarc.fr/ENG/Classification/index.php.
IARC. 2012. Agents classified by the IARC monographs. Lyon, France: World Health Organization,

International Agency for Research on Cancer Available from http://monographs.iarc.fr/ENG/Classification/index.php.
Jara A. 2007. Applied Bayesian non- and semi-parametric inference using DPpackage. R News 7(3): 17-26.
Jara A, Hanson T, Quintana FA, Müller P, Rosner GL. 2011. DPpackage: Bayesian semi- and nonparametric modeling in R. Journal of Statistical Software 40(5): 1-30.
Jean-Frédéric J, Gaël R, Thierry R. 2004. Financial applications of copula functions. Par Giorgio Szego: John Wiley and Sons.
Jenkinson AF. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society 81(348): 158-171.
Jia C, Batterman S, D'Souza J. 2010. Copulas and other multivariate models of personal exposures to VOC mixtures. Human & Ecological Risk Assessment.
Jia C, Batterman SA, Relyea GE. 2011. Variability of indoor and outdoor VOC measurements: An analysis using variance components. Environmental Pollution In press.
Jia C, D'Souza J, Batterman S. 2008. Distributions of personal VOC exposures: a population-based analysis. Environment International 34(7): 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. Science of The Total Environment 269(1-3): 25-37.
Kassomenos P, Lykoudis S, Chaloulakou A. 2010. A tool for determining urban emission characteristics to be used in exposure assessment. Environment International 36(3): 281-289.
Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. Advances in Water Resources 25(8): 1287-1304.
Kim H, Bernstein J. 2009. Air pollution and allergic disease. Current Allergy and Asthma Reports 9(2): 128-133.
Kim YM, Harrad S, Harrison RM. 2002. Levels and sources of personal inhalation exposure to volatile organic compounds. Environmental Science and Technology 36(24): 5405-5410.
Kinney P, Chillrud S, Ramstrom S, Ross J, Spengler J. 2002. Exposures to multiple air toxics in New York City. Environmental Health Perspectives 110(Suppl 4): 539-546.
Kinney PL, Chillrud SN, Ramstrom S, Ross J, Spengler JD. 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. Journal of Exposure Analysis and Environmental Epidemiology 11(3): 231-252.
Koenker R, Bassett G. 1978. Regression quantiles. Econometrica 46(1): 33-50.
Krishnamoorthy K, Mallick A, Mathew T. 2009. Model-based imputation approach for data analysis in the presence of non-detects. Annals of Occupational Hygiene 53(3): 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. Biological Research For Nursing 6(2): 151-157.
Kwok ESC, Atkinson R. 1995. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship: an update. Atmospheric Environment 29: 1685-1695.
Lenox MJ, Haimes YY. 1996. The constrained extremal distribution selection method. Risk Analysis 16(2): 161-176.
Lin YS, Egeghy PP, Rappaport SM. 2008. Relationships between levels of volatile organic compounds in air and blood from the general population. Journal of Exposure Science and Environmental Epidemiology 18(4): 421-429.
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.

Environmental Pollution 159(10): 2310-2319.
Lioy P, Wallace L, Pellizzari E. 1991. Indoor/outdoor, and personal monitor and breath analysis relationships for selected volatile organic compounds measured at three homes during New Jersey TEAM-1987. Journal of Exposure Analysis and Environmental Epidemiology 1(1): 45-61.
Lippy BE, Turner RW. 1991. Complex mixtures in industrial workspaces: Lessons for indoor air quality evaluations. Environmental Health Perspectives 95: 81-83.
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. Environmental Health Perspectives 112(17): 1691-1696.
McCarthy MC, Hafner HR, Chinkin LR, Charrier JG. 2007. Temporal variability of selected air toxics in the United States. Atmospheric Environment 41(34): 7180-7194.
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: Dekker, M.
McLachlan GJ, Krishnan T. 1997. The EM algorithm and extensions. New York, U.S.: Wiley & Sons.
McLachlan GJ, Peel D. 2000. Finite mixture models. New York, U.S.: Wiley and Sons.
MDE. 2010. Top ten sources of volatile organic compound in the Baltimore Area 2002. Maryland Department of the Environment. Available from http://www.mde.state.md.us/Air/air_information/TopTenVOC.asp.
Mendell MJ. 2007. Indoor residential chemical emissions as risk factors for respiratory and allergic effects in children: A review. Indoor Air 17(4): 259-277.
Meng XL, Pedlow S. 1992. EM: A bibliographic review with missing articles. The proceedings of the statistical computing section of the American Statistical Association: 24-27.
Mudarri DH. 2010. Building codes and indoor air quality. Available from http://wwwepagov/iaq/pdfs/building_codes_and_iaqpdf.
Mueller P, Quintana F. 2004. Nonparametric Bayesian data analysis. Statistical Science 19(1): 95-110.
NCHS. 2000. NHANES III priority toxicant reference range study data file. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHANES/NHANESIII/4A/voc-acc.pdf.
NCHS. 2006. The National Health and Nutrition Examination Survey (NHANES) analytic and reporting guidelines. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf.
NCHS. 2009. MEC interviewers procedures manual. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_09_10/MECInterviewers.pdf.
NCHS. 2010a. National Health and Nutrition Examination Survey 1999–2000 general data release documentation. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/gendoc.pdf.
NCHS. 2010b. National Health and Nutrition Examination Survey 2001-2002 general data release documentation. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_01_02/general_data_release_doc.pdf.
NCHS. 2010c. National Health and Nutrition Examination Survey 2003–2004 general data release documentation. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/general_data_release_doc_03-04.pdf.

NCHS. 2010d. National Health and Nutrition Examination Survey 1999–2000 data documentation, lab 04 volatile organic compounds in blood and water. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/frequency/lab04.pdf.
NCHS. 2010e. NHANES laboratory/medical technologists procedures manual. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from www.cdc.gov/nchs/data/nhanes/lab1-6.pdf.
NCHS. 2010g. Continuous NHANES web tutorial. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/tutorials/Nhanes/index_current.htm.
NCHS. 2011. Laboratory methodology and public data files. CDC National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD. Available from http://www.cdc.gov/nchs/data/nhanes/nhanes_99_00/lab_generaldoc.pdf.
NCHS. 2012a. National Health and Nutrition Examination Survey 1999-2000 Data Documentation, Codebook, and Frequencies: Volatile Organic Compounds (VOC) (LAB21) Hyattsville, MD: CDC National Center for Health Statistics, US Department of Health and Human Services, Centers for Disease Control and Prevention Available from http://www.cdc.gov/nchs/nhanes/nhanes1999-2000/LAB21.htm.
NCHS. 2012b. National Health and Nutrition Examination Survey 1999-2000 NHANES 1999 - 2000 Laboratory Files. Hyattsville, MD: US Department of Health and Human Services.
OEHHA. 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, Office of Environmental Health Hazard Assessment, 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.
Panko JM, Gaffney SH, Burns AM, Unice KM, Kreider ML, Booher LE, et al. 2009. Occupational exposure to benzene at the ExxonMobil Refinery at Baton Rouge, Louisiana (1977-2005). Journal of Occupational and Environmental Hygiene 6(9): 517 - 529.
Paulo MJ, van der Voet H, Wood JC, Marion GR, van Klaveren JD. 2006. Analysis of multivariate extreme intakes of food chemicals. Food and Chemical Toxicology 44(7): 994-1005.
Phillips M, Esmen N, Hall T, Lynch R. 2005. Determinants of exposure to volatile organic compounds in four Oklahoma cities. Journal of Exposure Analysis and Environmental Epidemiology 15(1): 35-46.
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(2): 519-527.
Redner R, Walker H. 1984. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review 26(2): 195-239.
Riederer A, Bartell S, Ryan P. 2009. Predictors of personal air concentrations of chloroform among US adults in NHANES 1999-2000. Journal of Exposure Science and Environmental Epidemiology 19(3): 248-259.
Rodgers JL, Nicewander WA. 1988. Thirteen ways to look at the correlation coefficient. The American Statistician 42(1): 59-66.
Rumchev K, Brown H, Spickett J. 2007. Volatile organic compounds: Do they present a risk to our health? Rev Environ Health 22(1): 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. Environmental Health Perspectives 114(10): 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, Adgate J, Church T, Ashley D, Needham L, Ramachandran G, et al. 2005. Children's exposure to volatile organic compounds as determined by longitudinal measurements in blood. Environ Health Perspect 113(3): 342-349.
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. Journal of Toxicology and Environmental Health, Part A 70(5): 465-476.
Sherlach KS, Gorka AP, Dantzler A, Roepe PD. 2011. Quantification of perchloroethylene residues in dry-cleaned fabrics. Environmental Toxicology and Chemistry 30(11): 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(3): 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 J. T. de Oliveira (ed.). Boston, MA: Statistical extremes and applications, (D. Reidel Publishing Co.).
Staudt A. 2010. Tail Risk, systemic risk and copulas. Casualty Actuarial Society E-Forum 2.
Stephens MA. 1974. EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association 69(347): 730-737.
Su F-C, Jia C, Batterman S. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. Atmospheric Environment 62(0): 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. Atmospheric Environment 45(28): 4858-4867.
Su F-C, Mukherjee B, Batterman S. 2013. Determinants of personal, indoor and outdoor voc concentrations: an analysis of the RIOPA data. In process.
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. Atmospheric Environment 21(8): 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. Journal of Toxicology and Environmental Health, Part A 72(14): 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(3): 681-688.
Titterington DM, Smith AFM, Makov UE. 1985. Statistical analysis of finite mixture distributions. Chichester, U.K: John Wiley and 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 and Chemical Toxicology 42(8): 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. Journal of Environmental Radioactivity 99(4): 649-657.
US EPA. 1986. Guidelines for the health risk assessment of chemical mixtures. Washington, DC: US Environmental Protection Agency Available from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=22567#Download.
US EPA. 1989. Report to Congress on indoor air quality: Volume 2. EPA/400/1-89/001C Washington,

DC: US Environmental Protection Agency.
US EPA. 1993. Provisional guidance for quantitative risk assessment of polycyclic aromatic hydrocarbons. Washington, DC: US Environmental Protection Agency Available from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=49732.
US EPA. 1996. National-Scale Air Toxics Assessment. Available from http://wwwepagov/ttn/atw/nata/.
US EPA. 1998. National volatile organic compound emission standards for architectural coatings Available from http://wwwepagov/EPA-AIR/1998/September/Day-11/a22659htm.
US EPA. 1999a. National-Scale Air Toxics Assessment. Available from http://wwwepagov/ttn/atw/nata1999/.
US EPA. 2000a. Supplementary guidance for conducting health risk assessment of chemical mixtures US Environmental Protection Agency, Washington, DC Available from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533.
US EPA. 2000b. Supplementary guidance for conducting health risk assessment of chemical mixtures Washington, DC: US Environmental Protection Agency Available from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533.
US EPA. 2000c. Taking toxics out of the air. Available from http://wwwepagov/air/oaqps/takingtoxics/indexhtml.
US EPA. 2001. Urban Air Toxics Monitoring Program. Available from http://wwwepagov/ttnamti1/uatmhtml.
US EPA. 2002. National-Scale Air Toxics Assessment. Available from http://wwwepagov/ttn/atw/nata2002/.
US EPA. 2003. Framework for cumulative risk assessment. Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/raf/publications/pdfs/frmwrk_cum_risk_assmnt.pdf.
US EPA. 2003b. Toxicological review of xylenes. Available from http://wwwepagov/iris/toxreviews/0270trpdf.
US EPA. 2007a. Mobile source air toxics - regulations. Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/oms/toxics-regs.htm.
US EPA. 2007b. Trichloroethylene (TCE) TEACH chemical summary. Available from http://wwwepagov/teach/chem_summ/TCE_summarypdf.
US EPA. 2008a. EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. Available from http://wwwepagov/heasd/products/pmf/EPA%20PMF%2030%20User%20Guide%20v16_092208_finalpdf.
US EPA. 2008b. Regulatory determinations support document for selected contaminants from the second drinking water contaminant candidate list (CCL 2). Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/ogwdw/ccl/pdfs/reg_determine2/report_ccl2-reg2_supportdocument_full.pdf.
US EPA. 2009. Integrated Risk Information System (IRIS) glossary/acronyms and abbreviations. Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/iris/help_gloss.htm.
US EPA. 2009a. Ambient concentrations of benzene Available from http://cfpubepagov/eroe/indexcfm?fuseaction=detailviewInd&lv=listlistByAlpha&r=216617&subtop=341.
US EPA. 2009b. Air toxics emissions Available from http://cfpubepagov/eroe/indexcfm?fuseaction=detailviewInd&ch=46&subtop=341&lv=listlistByChapter&r=216616.
US EPA. 2010a. Comparison of 1999 model-predicted concentrations to monitored data. Available from http://wwwepagov/ttn/atw/nata1999/99comparehtml.

US EPA. 2010b. Control emmissions technologies - transport & dispersion of air pollutants. Available from http://wwwepagov/apti/course422/ce1html.
US EPA. 2010c. Mobile source air toxics Available from http://wwwepagov/oms/toxicshtm.
US EPA. 2010b. Volatile organic compounds emissions Available from http://cfpubepagov/eroe/indexcfm?fuseaction=detailviewInd&lv=listlistByAlpha&r=209842&subtop=341.
US EPA. 2010c. Summaries of EPA's final air toxics MACT rules. Available from http://wwwepagov/air/oaqps/takingtoxics/p2html
US EPA. 2011a. Air Quality System (AQS). Available from http://wwwepagov/ttn/airs/airsaqs/detaildata/downloadaqsdatahtm.
US EPA. 2011b. Indoor air. Available from http://cfpubepagov/eroe/indexcfm?fuseaction=listlistBySubTopic&ch=46&s=343.
US EPA. 2012a. Integrated Risk Information System (IRIS). Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/IRIS/index.html.
US EPA. 2012b. An introduction to indoor air quality: Volatile organic compounds (VOCs). Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/iaq/voc.html.
US EPA. 2012c. MTBE- recommendations and actions. Washington, DC: US Environmental Protection Agency Available from http://www.epa.gov/mtbe/action.htm.
US EPA. 2012d. National Ambient Air Quality Standards (NAAQS). Available from http://wwwepagov/air/criteriahtml.
US EPA. 2013. Basic information about disinfection byproducts in drinking water: total trihalomethanes, haloacetic acids, bromate, and chlorite. Available from http://waterepagov/drink/contaminants/basicinformation/disinfectionbyproductscfm.
Wallace L, Pellizzari E, Hartwell TD, Perritt R, Ziegenfus R. 1987. Exposures to benzene and other volatile compounds from active and passive smoking. Archives of Environmental Health: An International Journal 42(5): 272 - 279.
Wallace LA. 2001. Human exposure to volatile organic pollutants: Implications for indoor air studies. Annual Review of Energy and the Environment 26(1): 269-301.
Wallace LA, Pellizzari E, Leaderer B, Zelon H, Sheldon L. 1987. Emissions of volatile organic compounds from building materials and consumer products. Atmospheric Environment 21(2): 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. Environmental Research 50(1): 37-55.
Wang B, Takigawa T, Takeuchi A, Yamasaki Y, Kataoka H, Wang D, et al. 2007. Unmetabolized VOCs in urine as biomarkers of low level exposure in indoor environments. Journal of Occupational Health 49(2): 104-110.
Wang S-W, Majeed MA, Chu P-L, Lin H-C. 2009. Characterizing relationships between personal exposures to VOCs and socioeconomic, demographic, behavioral variables. Atmospheric Environment 43(14): 2296-2302.
Weibull W. 1951. A statistical distribution function of wide applicability. Journal of Applied Mechanics 18(3): 293-297.
Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005a. Relationship of Indoor, Outdoor and Personal Air (RIOPA) study: Study design, methods and quality assurance/control results. Journal of Exposure Analysis and Environmental Epidemiology 15(2): 123-137.
Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005b. Relationships of Indoor, Outdoor, and Personal Air (RIOPA). Part I. Collection methods and descriptive analyses. Health Effects Institute Research Report (130 Pt 1): 1-127.
Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005c. Relationships of

Indoor, Outdoor, and Personal Air (RIOPA). Part I. Collection methods and descriptive analyses. Res Rep Health Eff Inst(130 Pt 1): 1-107; discussion 109-127.
Weschler CJ. 2011. Chemistry in indoor environments: 20 years of research. Indoor Air 21(3): 205-218.
WHO. 2012. The determinants of health. Geneva, Switzerland: UN World Health Organization Available from http://www.who.int/hia/evidence/doh/en/.
WHO. 2013. Metrics: Population Attributable Fraction (PAF). Geneva, Switzerland: UN World Health Organization Available from http://www.who.int/healthinfo/global_burden_disease/metrics_paf/en/index.html.
Wolkoff P, Schneider T, Kildesø J, Degerth R, Jaroszewski M, Schunk H. 1998. Risk in cleaning: chemical and physical exposure. Science of The Total Environment 215(1-2): 135-156.
Wu, Y., 1996. An application of hierarchical linear models to longitudinal studies. Research in Nursing and Health. 19, 75-82.