APPENDIX AVAILABLE ON THE HEI WEB SITE

Research Report 181

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

Stuart Batterman et al.

Appendix A. Supplemental Information

Correspondence may be addressed to Dr. Stuart Batterman, 1420 Washington Heights, Ann Arbor, MI 48109-2029;
e-mail: *stuartb@umich.edu*

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

This document was reviewed by the HEI Health Review Committee but did not undergo the HEI scientific editing and production process.

APPENDIX A. SUPPLEMENTAL INFORMATION

Table A.1 shows the determinants of 15 personal exposures to volatile organic compounds (VOCs) identified in 12 previous studies and in the present study. These determinants were grouped into three categories: personal activities, socioeconomic factors, and environmental factors.

Table A.2 summarizes cancer and non-cancer toxicity for VOCs measured in the Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study based on selected standards and guidelines. For cancer toxicity, Integrated Risk Information System (IRIS) and International Agency for Research on Cancer (IARC) cancer classifications and inhalation unit risk factors (URF) are shown. For non-cancer toxicity, chronic reference concentration (RfC) and acute minimal risk levels are presented.

Table A.1. Determinants of VOC exposures in previous studies and this study.

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 environmental tobacco smoke	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 or 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															
Air exchange rate		m	m	m	m					m	m		m	m	m
Ambient relative humidity										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 commercial 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										

An * indicates no increasing or decreasing trend; capital letters indicate increased exposure and lower case letters indicate decreased exposure from these studies: 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.

Table A.2. Summary of VOC toxicity, including selected standards and guidelines.

VOCs	Cancer					Non-cancer			
	IRIS	IARC	URF (per μg/m³)	Source	Health effect	Chronic RfC (μg\m³)	Source	Acute MRL (μg/m³)	Source
Benzene	A	1	7.8×10^{-6}	IARC 2014, US EPA 2014	Leukemia (occupational)	30	US EPA 2014	29	ATSDR 2013
Toluene	D	3	NA	IARC 2014, US EPA 2014	Neurological effects (occupational); color vision impairment (occupational) and respiratory irritation (human volunteer)	5000	US EPA 2014	3766	ATSDR 2013
Ethylbenzene	D	2B	2.5×10^{-6}	IARC 2014, OEHHA 2005	Lung, liver, and renal adenomas and carcinomas (animal)	1000	US EPA 2014	21696	ATSDR 2013
Xylenes	D	3	NA	IARC 2014, US EPA 2014	Impaired motor coordination (animal)	100	US EPA 2014	8679	ATSDR 2013
MTBE	D	3	2.6×10^{-7}	IARC 2014, OEHHA 2005	Lymphomas, leukaemias, hepatocellular adenomas, and renal tubular and testicular tumours (animal)	3000	US EPA 2014	7206	ATSDR 2013
Styrene	ND	2B	2.0×10^{-6}	IARC 2014, Caldwell et al. 1998	Pulmonary adenomas (animal)	1000	US EPA 2014	21286	ATSDR 2013
1,4-DCB	ND	2B	1.1×10^{-5}	IARC 2014, OEHHA 2005	Liver and kidney tumor, and mononuclear-cell leukemia (animal)	800	US EPA 2014	12019	ATSDR 2013
TCE	ND	2A	2.0×10^{-6}	IARC 2014, OEHHA 2005	Liver and biliary tract cancer, and lymphoma (human); liver, renal-cell, lung and testicular tumours, and lymphomas (animal)	40	US EPA 2001	10741	ATSDR 2013
PERC	ND	2A	5.9×10^{-6}	IARC 2014, OEHHA 2005	Oesophageal and cervical cancer, and non-Hodgkin's lymphoma (human); hepatocellular carcinomas and mononuclear-cell leukaemia (animal)	16	US EPA 2010	1356	ATSDR 2013
Chloroform	B2	2B	2.3×10^{-5}	IARC 2014, US EPA, 2014	Renal tubule and hepatocellular tumours (animal)	NA		488	ATSDR 2013
CTC	B2	2B	1.5×10^{-5}	IARC 2014, US EPA 2014	Liver and mammary neoplasms (animal)	100	US EPA 2014	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; ATSDR, Agency for Toxic Substance and Disease Registry; OEHHA, Office of Environmental Health Hazard Assessment. US EPA, US Environmental Protection Agency; URF, unit risk factor; RfC, reference concentration; MRL, minimal risk level; NA, not available; ND, no data.

For IRIS cancer classifications, A indicates a human carcinogen, B2 indicates a probable human carcinogen (based on sufficient evidence of carcinogenicity in animals), and D indicates not classifiable as to human carcinogenicity.

For IARC cancer classifications, 1 indicates carcinogenic to humans, 2A indicates probably carcinogenic to humans, 2B indicates possibly carcinogenic to humans, and 3 indicates not classifiable as to its carcinogenicity to humans.

1. 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. For each VOC measurement type (outdoor, indoor, adult personal, child personal), all observations (i.e., samples from both first and second visits) before and after log transformation were used for full distribution fitting.

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

Table A.3 shows the distribution types providing the highest GOF, based on A-D tests, by VOC measurement type (outdoor, indoor, adult personal, child personal). 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 A.1 through A.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 shows a large number of low-concentration measurements; this is a result of including concentrations below the MDL, which are typically addressed by setting values to one-half MDL or some similar value. The detection frequencies (percentage of values above the MDL) ranged from 6% to 97% for outdoor measurements; 25% to 95% for indoor measurements; 31% to 96% for adult personal measurements; and 23% to 97% for child personal measurements. Outdoor VOCs, (including styrene, 1,4-DCB, MC, TCE, chloroform, *d*-limonene, α-pinene, and β-pinene) plus 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 A.1 through A.4 also point out positive skew after log transformation and outliers that remain high and 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 great interest given that these portray the highest exposures.

The full-range distributions of VOCs in RIOPA and NHANES shared some similarities. For the NHANES VOCs, distributions were right-skewed and the top-ranked distributions were usually lognormal (except for MTBE, 1,4-DCB, and TCE). In contrast, for 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 A.5 contrasts observed and modeled distributions for benzene, 1,4-DCB, PERC, and chloroform, which can be compared with the panels for adult personal distributions shown in Figures A.1 through A.4. This analysis showed a number of differences. First, 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 significantly different (Mann-Whitney tests, $P < 0.05$); average concentrations were not different for only three compounds (1,4-DCB, PERC, and chloroform, t test, $P < 0.05$).

Table A.3. Identification of best-fit distributions (top ranked) for VOCs by sample type.

VOCs	Best-fit distribution								
	Untransformed					Log-transformed			
	Outdoor (n = 555)	Indoor (n = 554)	Adult (n = 544)	Adult_NH (n = 665)	Child (n = 209)	Outdoor (n = 555)	Indoor (n = 554)	Adult (n = 544)	Child (n = 209)
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
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. The RIOPA dataset was used for all other analyses,.

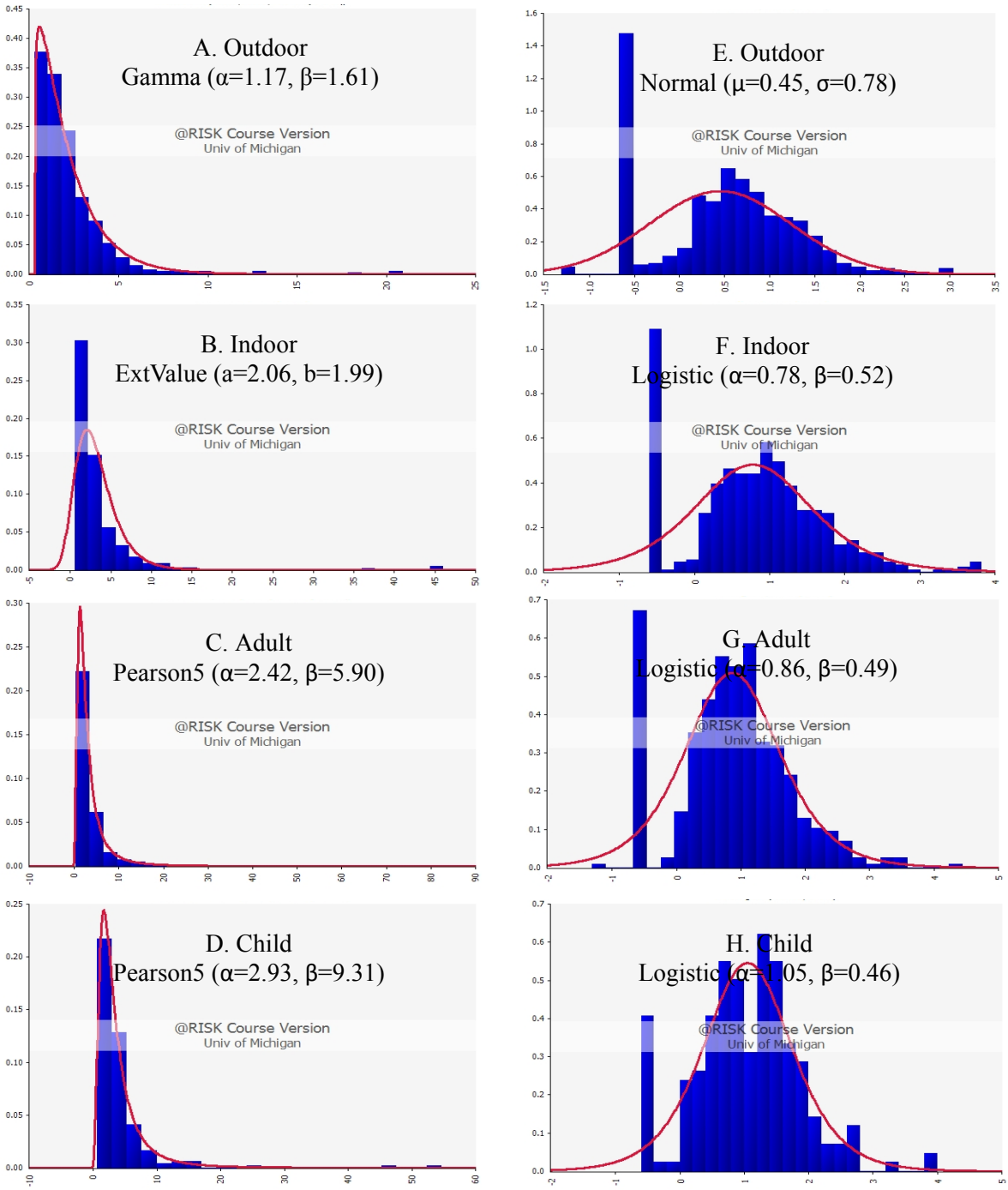

Figure A.1. Histograms of observed and fitted distributions of benzene by sample type. Observed shown as blue bars; fitted shown as red line; left panels (A–D) are untransformed data; right panels (E–H) use natural log transform.

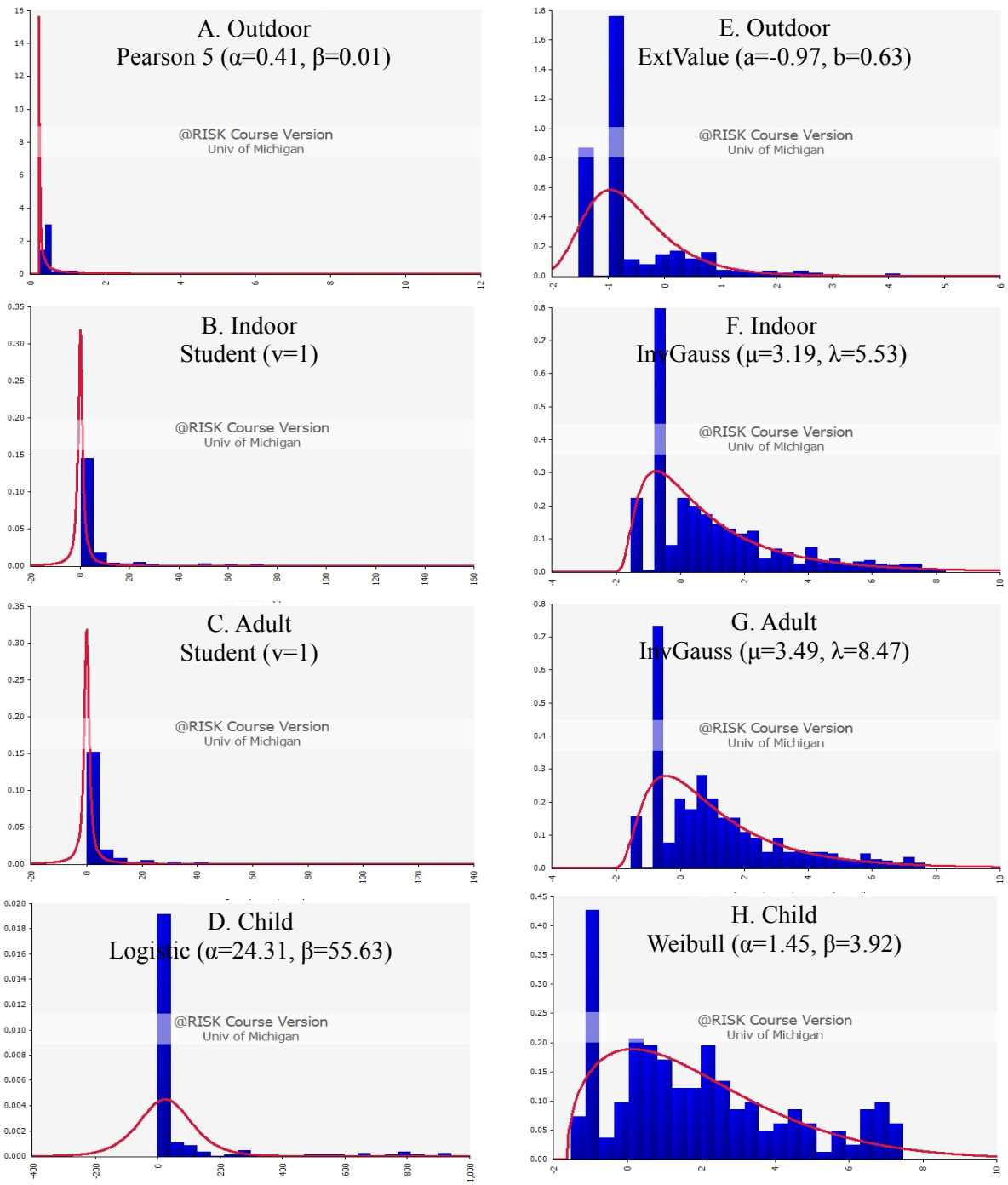

Figure A.2. Histograms of observed and fitted distributions of 1,4-DCB by sample type. Observed shown as blue bars; fitted shown as red line; left panels (A–D) are untransformed data; right panels (E–H) use natural log transform. Plots omit some data: 1,4-DCB concentrations > 5 μg/m³ (n = 23), 150 μg/m³ (n = 41), 150 μg/m³ (n = 38), and 1000 μg/m³ (n = 10) in panels A, B, C and D, respectively.

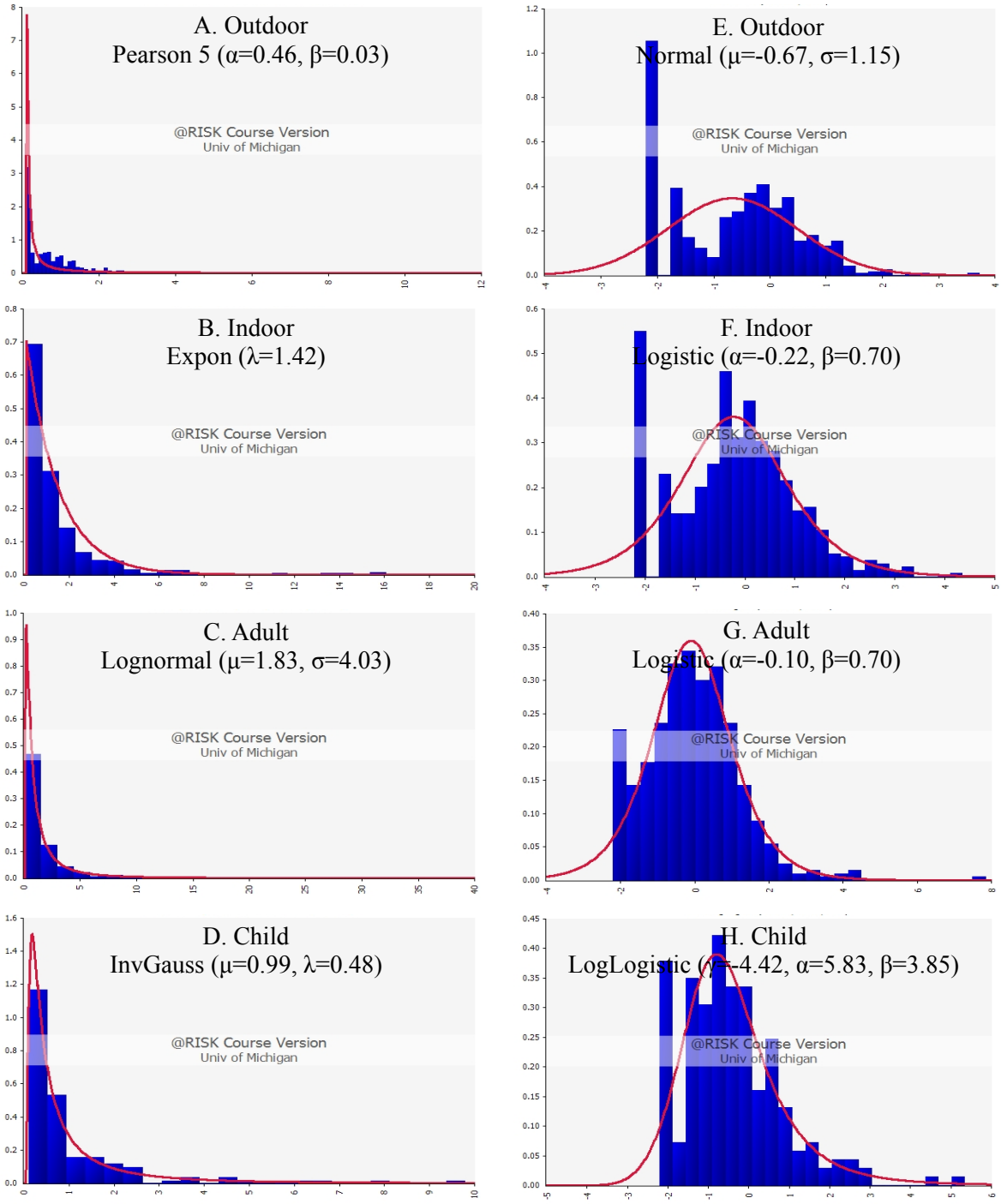

Figure A.3. Histograms of observed and fitted distributions of PERC by sample type. Observed shown as blue bars; fitted shown as red line; left panels (A–D) are untransformed data; right panels (E–H) use natural log transform. Plots omit some data: PERC concentrations > 3 μg/m^3 (n = 32), 30 μg/m^3 (n = 1), 40 μg/m^3 (n = 6), and 20 μg/m^3 (n = 2) in panels A, B, C and D, respectively.

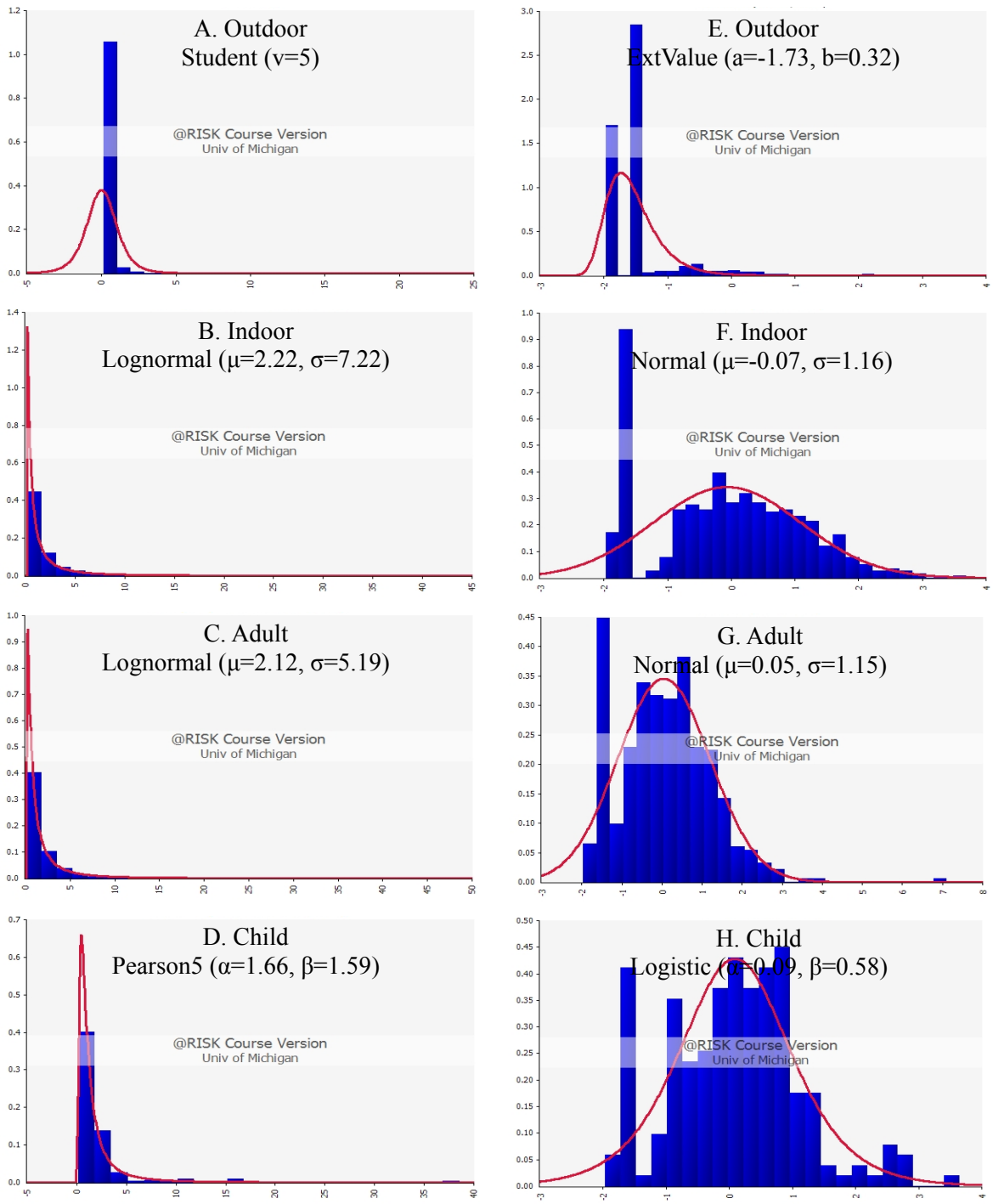

Figure A.4. Histograms of observed and fitted distributions of chloroform by sample type. Observed shown as blue bars; fitted shown as red line; left panels (A–D) are untransformed data; right panels (E–H) use natural log transform. Plots omit one data point: chloroform concentrations > 1224 μg/m^3 (n = 1) in panel C.

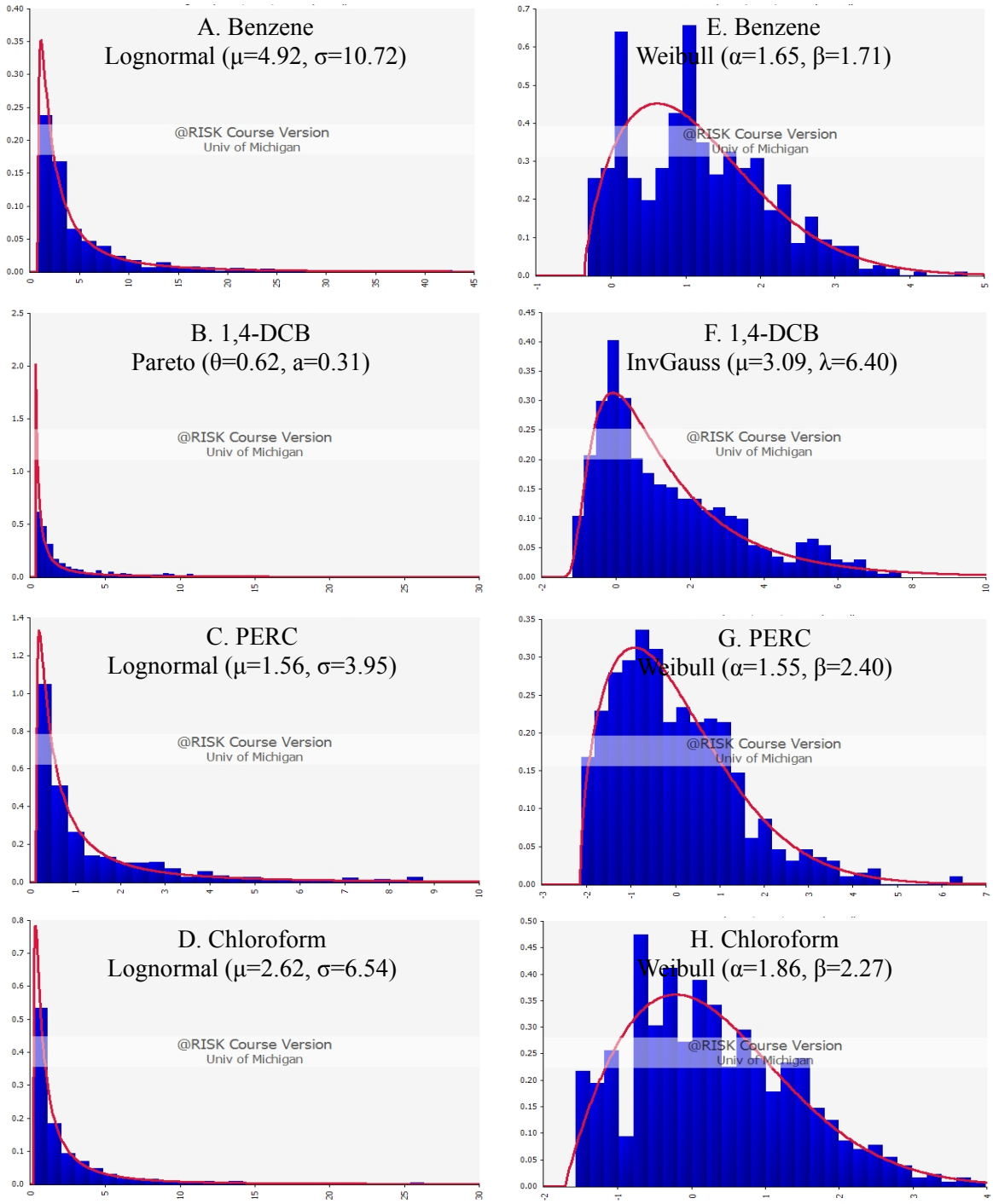

Figure A.5. Histograms of observed and fitted distributions of benzene, 1,4-DCB, PERC, and chloroform concentrations in 1999/2000 NHANES. Observed shown as blue bars; fitted shown as red line; left panels (A–D) are untransformed data; right panels (E–H) use natural log transform. Plots omit some data: 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).

2. Gumbel Distribution Fitting

The main report describes fitting using generalized extreme value (GEV) distributions. This section discusses Gumbel distributions, a two-parameter form of GEV, that are fitted using probability plots, a different approach from that used in the report.

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 personal 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 (Barnett 1975). First, extrema were ranked in descending order. Then, each observation was plotted against $-\ln[-\ln(Pv)]$, where Pv (estimated probability of each observation, v) was computed as:

$$Pv = (r - 0.44)/(n + 0.12) \quad (1)$$

where r is the reverse rank of VOC concentrations, and n is 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.

Figures A.6 through A.9 display model fits to the data for indoor, outdoor, and personal (adult and child) concentrations for the same four VOCs discussed earlier (benzene, 1,4-DCB, PERC, and chloroform). Table A.4 summarizes results for all VOCs and sample types.

- In all cases, Gumbel distributions provided a better fit to extrema defined as values above the 95th percentile, as compared with values above the 90th percentile, suggesting that the 95th percentile 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; adult personal measurements of ethylbenzene, *m*- & *p*-xylene, *o*-xylene, styrene, 1,4-DCB, and β-pinene; and child personal 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, 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, including PERC and CTC.
- In a number of cases, an even higher cut-off might have been appropriate when fitting Gumbel-type distributions, and sometimes results were driven by a few outliers.

Overall, these results suggest that simple parametric distributions (e.g., lognormal distributions) do not fit the entire range of observations in the RIOPA VOC dataset; extreme value distributions often can provide good fits for 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 suggest that extreme values are more likely for certain VOCs and certain types of exposure measures (for example, high personal exposures to BTEX may be associated with vehicle refueling events and high indoor levels of pinenes may be associated with cleaning events); and that for some VOCs and certain exposure compartments (microenvironments), outliers are unlikely (for example, CTC is a long-lived VOC with few localized sources, and other solvents and some other VOCs also have only a few strong and localized outdoor sources likely to produce extrema).

Gumbel distributions for NHANES have been shown in a previous study (Jia et al. 2008). There are several differences in comparison with distributions determined for the RIOPA dataset. Most VOCs in NHANES had better fits (higher R^2) to the maximum Gumbel distribution than did the RIOPA data, although BTEX compounds had high R^2 values in both data sets. Chlorinated hydrocarbons (TCE, PERC, and chloroform) had better fits in the NHANES data set (compared with RIOPA), but worse fits for 1,4-DCB. Several large differences were seen in maxima in that RIOPA had higher maximum concentrations, sometimes by very large amounts; for example, PERC and chloroform maxima in RIOPA were 2,618 and 1,224 $\mu g/m^3$, respectively, compared with 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 in the RIOPA data.

Different sampling designs and sample biases likely explain some of the differences between RIOPA and NHANES. First, since NHANES was designed as a nationally representative sample, that data set should reflect population heterogeneity; and if this applies to VOCs and extrema, then NHANES should better represent the true extreme value distributions than would the more stratified sampling design used in RIOPA. Second, study protocols differed in important ways. In NHANES, staging was extensive; it included two trips by participants, in most cases by private vehicle, to a centrally located mobile examination center (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. The common staging might have produced greater uniformity in the NHANES data. However, 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 in personal air in a subset of the 1999–2000 NHANES cohort (Su et al. 2012). These discrepancies, however, are not expected to adversely affect the comparability of the air samples between the two studies.

Table A.4. Goodness of fit measures (R^2) for the maximum Gumbel distribution fits to 90th and 95th extrema 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**
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**

NA, not available; adult_NH, personal airborne exposures in the 1999/2000 NHANES database. The RIOPA dataset was used .for all other analyses.

$R^2 < 0.6$ are shown in red, and $R^2 > 0.85$ are shown in blue bold type.

Tables A.5 through A.9 are supplemental materials for the section on extreme value analyses in the main text.

Table A.5. GEV parameters and goodness-of-fit (*P* value) for average VOC exposures (μg/m^3) 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	0.763	-0.2	13.6	3.6	0.417
Toluene	1.6	35.8	7.3	0.435	0.6	63.6	19.2	0.837
Ethylbenzene	1.2	6.3	1.7	0.991	0.8	10.6	3.9	0.962
m-&*p*-Xylene	0.8	19.9	6.6	0.978	1.2	28.7	6.9	0.770
o-Xylene	0.9	6.8	2.1	0.438	1.8	10.0	1.3	0.755
MTBE	0.6	36.3	12.5	0.980	0.9	53.0	11.4	0.946
Styrene	1.3	3.9	1.6	0.402	0.9	8.4	2.8	0.894
1,4-DCB	0.5	258.0	188.0	0.966	0.5	516.0	234.9	0.927
TCE	1.1	1.7	0.8	0.973	1.7	2.8	1.0	0.733
PERC	1.0	5.9	2.6	0.906	0.7	11.4	4.2	0.994
Chloroform	0.7	5.5	1.6	0.963	1.1	7.6	1.7	0.853
CTC	0.7	0.9	0.1	0.792	0.7	1.1	0.1	0.906
d-Limonene	0.6	85.8	20.0	0.611	0.4	124.8	19.7	0.783
α-Pinene	1.1	18.0	4.0	0.810	1.7	23.4	6.0	0.799
β-Pinene	0.9	18.2	6.5	0.630	0.1	35.2	13.8	0.815

* *P* values shown for Kolmogorov–Smirnov tests; *P* values > 0.05 indicate that observations fit to generalized extreme value distributions.

Table A.6. GEV parameters and goodness-of-fit for personal VOC exposures (μg/m^3) in NHANES using weighted dataset.

VOCs	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 indicates that observations fit to GEV distributions or that the observations were not different from GEV simulations (one value in bold).

Table A.7. GEV parameters and goodness-of-fit for personal VOC exposures (μg/m^3) in NHANES using bootstrap methods and repeated sampling.

VOCs	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 indicates that observations fit to GEV distributions, or that the observations were not different from GEV simulations (two values in bold).

Table A.8. GEV parameters and goodness-of-fit for the personal VOC exposures (μg/m^3) in NHANES using unweighted dataset.

VOCs	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.

All P values are > 0.05, which indicates that observations fit to GEV distributions or that the observations were not different from GEV simulations.

Table A.9. Posterior distribution of the number of clusters K for several prior settings of α.

	Posterior distribution of K								
	Chloroform (n = 544)			1,4-DCB (n = 544)			Styrene (n = 544)		
Prior	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: $\alpha \sim \text{Gamma}(5, 2)$

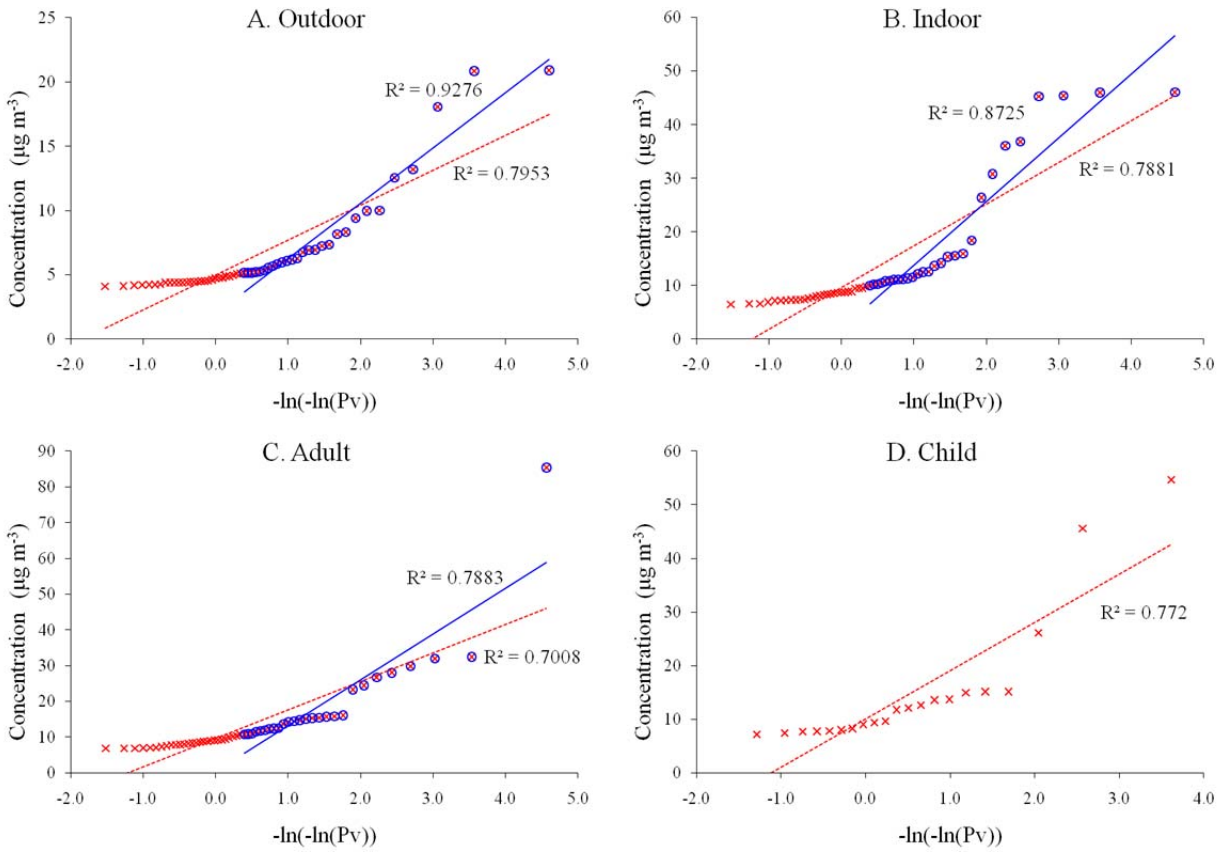

Figure A.6. Probability plots for Gumbel distributions of benzene. Top 10% (red x and dashed line, n = 56, 56, 54, and 21 for outdoor, indoor, adult, and child) and 5% (blue circle and solid line, n = 28, 28, and 27 for outdoor, indoor, and adult) of benzene concentrations fitted to maximum extreme distributions by sample type. $Pv = (r - 0.44)/(n + 0.12)$, where r = the reverse rank of Ci (concentration of each observation), and n = number of the extreme values (Barnett, 1975).

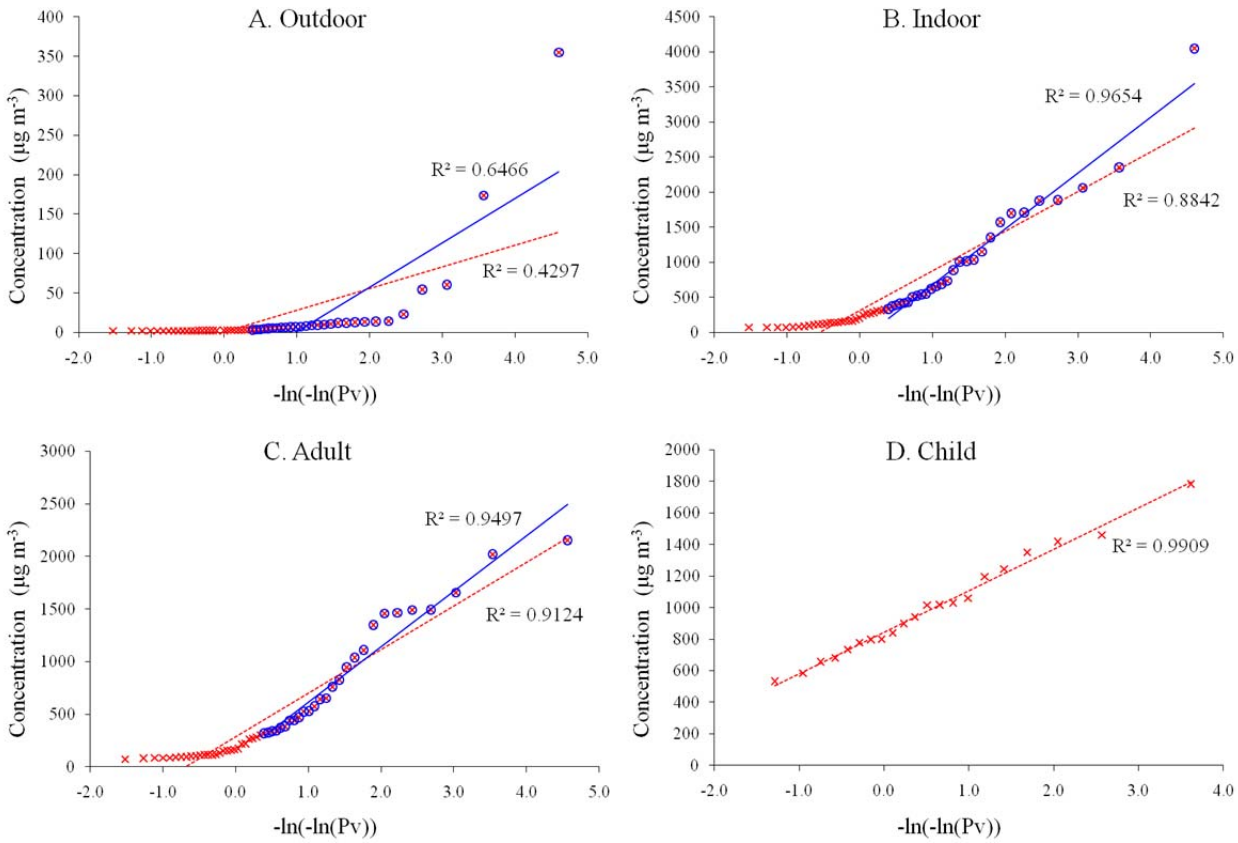

Figure A.7. Probability plots for Gumbel distributions of 1,4-DCB. Top 10% (red x and dashed line, n = 56, 56, 54, and 21 for outdoor, indoor, adult, and child) and 5% (blue circle and solid line, n = 28, 28, and 27 for outdoor, indoor, and adult) of 1,4-DCB concentrations fitted to maximum extreme distributions 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).

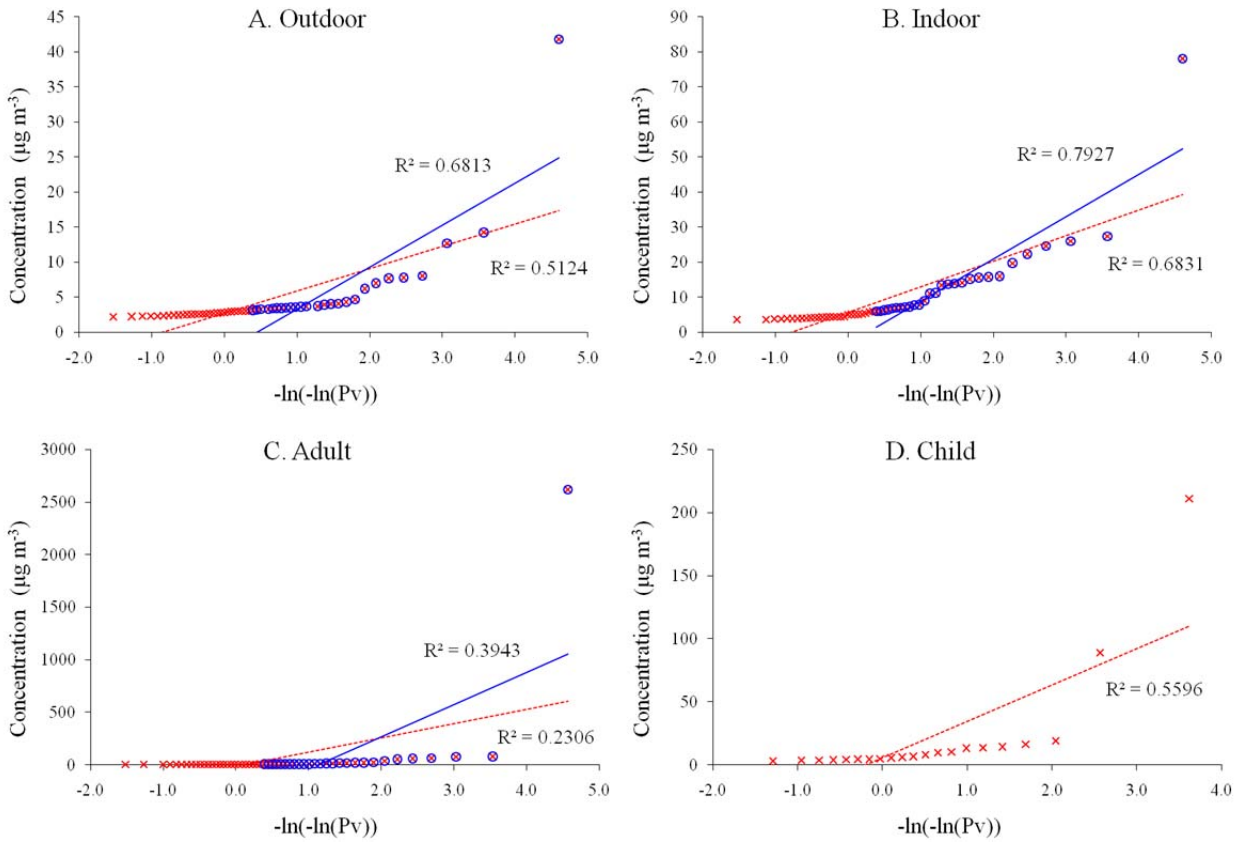

Figure A.8. Probability plots for Gumbel distributions of PERC. Top 10% (red x and dashed line, n = 56, 56, 54, and 21 for outdoor, indoor, adult, and child) and 5% (blue circle and solid line, n = 28, 28, and 27 for outdoor, indoor, and adult) of PERC concentrations fitted to maximum extreme distributions 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).

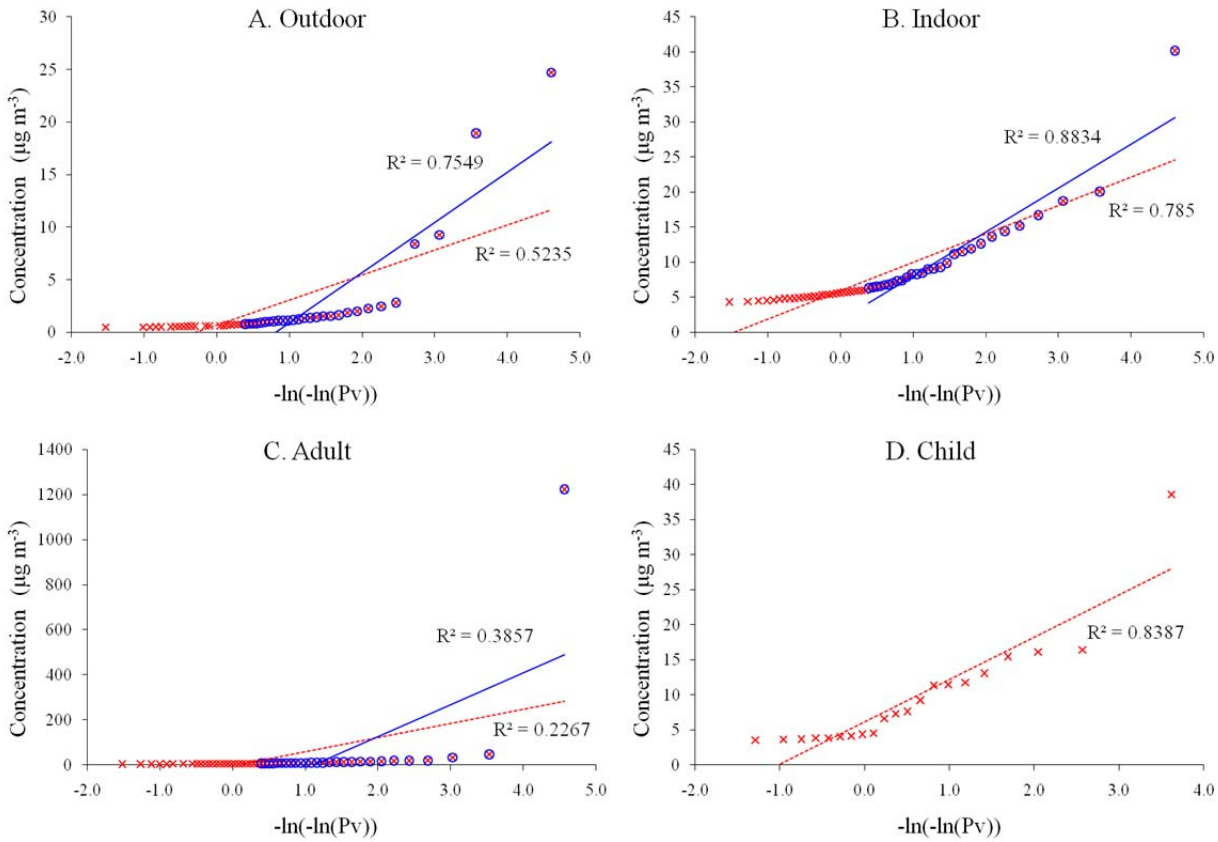

Figure A.9. Probability plots for Gumbel distributions of chloroform. Top 10% (red x and dashed line, n = 56, 56, 54, and 21 for outdoor, indoor, adult, and child) and 5% (blue circle and solid line, n = 28, 28, and 27 for outdoor, indoor, and adult) of chloroform concentrations fitted to maximum extreme 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).

3. **Selected VOC Mixtures Based on PMF**

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; the final group (presented in the main section of this Research Report) 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; for example, children would not be expected to have occupational exposures.

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

3.1. Outdoor VOCs

Outdoors, apportionments were dominated by gasoline-related sources, and seasonal variation was observed. Results of source apportionment of VOCs in the RIOPA study are presented in Appendix Table A.10. 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 also 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; mixture 3 included a lot of VOCs (e.g., 1,4-DCB, TCE, CTC, *d*-limonene, α-pinene, and β-pinene), which may come from industrial emissions; in mixture 4 PERC was the dominant VOC and is used in the dry cleaning industry. Gasoline-related sources (more than 60% of the contributions) were prevailing for outdoor VOCs in both seasons.

Appendix Figure A.10 presents the median ratios of four common VOC groups — aromatics, MTBE, chlorocarbons, and terpenes — by quintiles of total VOC (TVOC) concentrations in order 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. MTBE was more abundant in the middle and highest quintiles. The gasoline-related VOCs (displayed as aromatics and MTBE) 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 were terpenes; the abundance dropped to 5% in the last quintile. Outdoor terpenes are emitted from biogenic sources. Higher concentrations of TVOC may be attributable to other VOCs primarily from anthropogenic sources. Thus, the abundance of terpenes decreased in the highest quintiles due to increases of other VOC concentrations. VOC measurements in different cities and seasons showed similar abundances with overall measurements, except for samples in Houston, which had a greater abundance of MTBE in the higher quintiles.

3.2. 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 (Appendix Table A.10): 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 used for degreasing; and mixture 4 was dominated by MTBE, which indicates gasoline vapor. Cleaning products and odorants were the leading emission sources for indoor VOCs in both warm (73% of the contribution) and cold (66%) seasons.

Aromatics and MTBE showed lower abundances at higher quintiles for indoor VOCs (Appendix Figure A.11). The abundance of gasoline-related VOCs (displayed as aromatics and MTBE) in the 5th quintile was about 16% compared with 44% in the 1st quintile, and there was no difference between warm and cold seasons. Although indoor gasoline-related VOCs are mainly generated by outdoor sources, indoor concentrations can be affected by infiltration (or air exchange) rates and indoor lifetimes. Other VOCs (e.g., 1,4-DCB and *d*-limonene) that are generated by indoor sources had extreme values that led to high abundance in the higher quintiles. For example, the average concentration of 1,4-DCB in the 4th quintile of TVOC was 10 $\mu g/m^3$; this represents a median abundance of only 1.8%; the average in the 5th quintile was 327 $\mu g/m^3$, which gave a much higher median abundance of 27%. Similar patterns were observed for *d*-limonene. Variations of VOC abundance were shown among cities, especially in Houston, where a quarter of 1,4-DCB samples in the 5th quintile were above 1,000 $\mu g/m^3$. In contrast, only one 1,4-DCB sample was above 1,000 $\mu g/m^3$ in Los Angeles and Elizabeth.

3.3. Personal VOCs Consisting of Adult and Child Measurements

Dominant VOC sources for personal exposures were cleaning products and odorants; seasonal effects were also observed (Appendix Table A.10). In the warm season, four groups of VOCs were shown: mixture 1 included *d*-limonene, α-pinene, and β-pinene, indicating the use of cleaning products and odorants; mixture 2 included ethylbenzene, *m*- & *p*-xylene, and *o*-

xylene, representing vehicle exhaust; in mixture 3, benzene and MTBE indicated gasoline vapor; and mixture 4 (containing 1,4-DCB, TCE, PERC, chloroform, and CTC) suggested exposures to moth repellents and chlorinated solvents. In the cold season, VOC apportionments were still dominated by cleaning products and odorants, like *d*-limonene, α-pinene and β-pinene (mixture 1; 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 (displayed as aromatics and MTBE) were less abundant at higher quintiles with some variation between cities (Appendix Figure A.12). Personal samples showed more abundance of chlorocarbons (TCE, PERC, chloroform, 1,4-DCB) in the highest quintile than did the 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 $\mu g/m^3$ for indoor samples, and 95 $\mu g/m^3$ for personal samples. No significant differences in abundance were found between seasons. However, large variations were observed among cities, especially in Houston. Chlorocarbons were the majority (85%) in the highest quintile in Houston and other VOC groups were below 10%. On the other hand, aromatics and terpenes were dominant in the highest quintiles in Los Angeles and Elizabeth. Most of the extreme values of 1,4-DCB were measured in Houston; for example, 18 of 66 1,4-DCB measurements were above 1,000 $\mu g/m^3$. In contrast, only two measurements in Elizabeth exceeded this 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.

3.4. 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 with those obtained using the original data (US EPA 2008). The variability of the results using the bootstrap samples shows the stability of original results. We used 500 runs, a sample size approximately the same as in RIOPA, random sampling with replacement, and personal VOC exposures. Appendix Figure A.13 represents the variability for each species of the profiles using box plots. The original results are shown (as blue boxes) for reference. Although 2 to 4 of the VOCs in each factor had 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 was small, and the original results were consistent with the medians of the bootstrap model results. By this evaluation, source apportionment results using PMF were robust.

Table A.10. PMF results for outdoor, indoor, and personal VOCs in RIOPA. Suggested sources and apportionments are shown for (un-averaged measurements) by sample type and seasons (n = 555 for outdoor; n = 554 for indoor; n = 544 for personal).

Type	Season	Mixture	Suggested Source Categories	VOC Components	Fraction of TVOC	
					%	μg/m³
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.

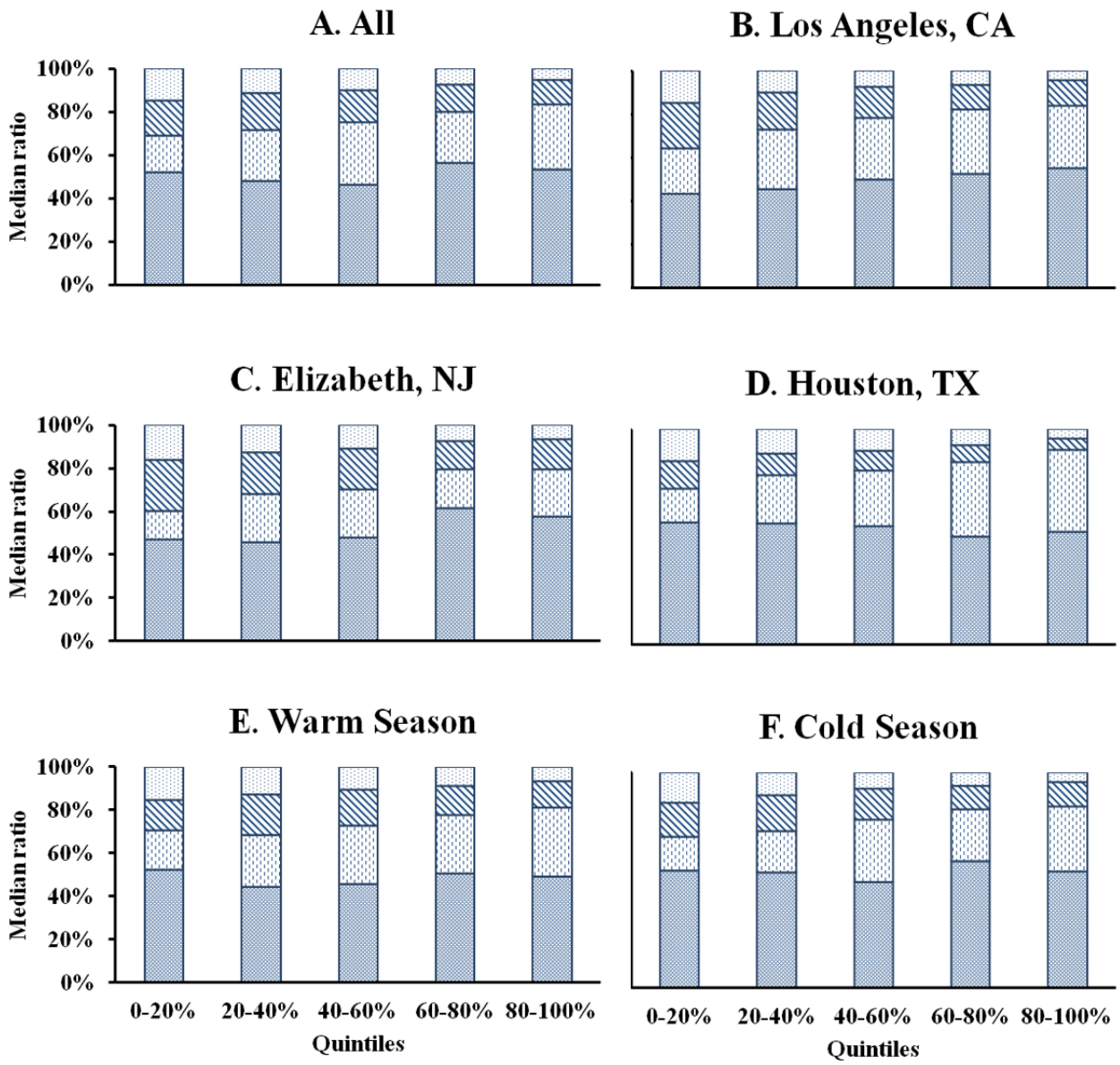

Figure A.10. Outdoor VOC composition at quintiles of total VOC concentrations. Warm season indicates April to September, and cold season indicates October to March (n = 555).
, aromatics; , MTBE; , chlorocarbons; , terpenes.

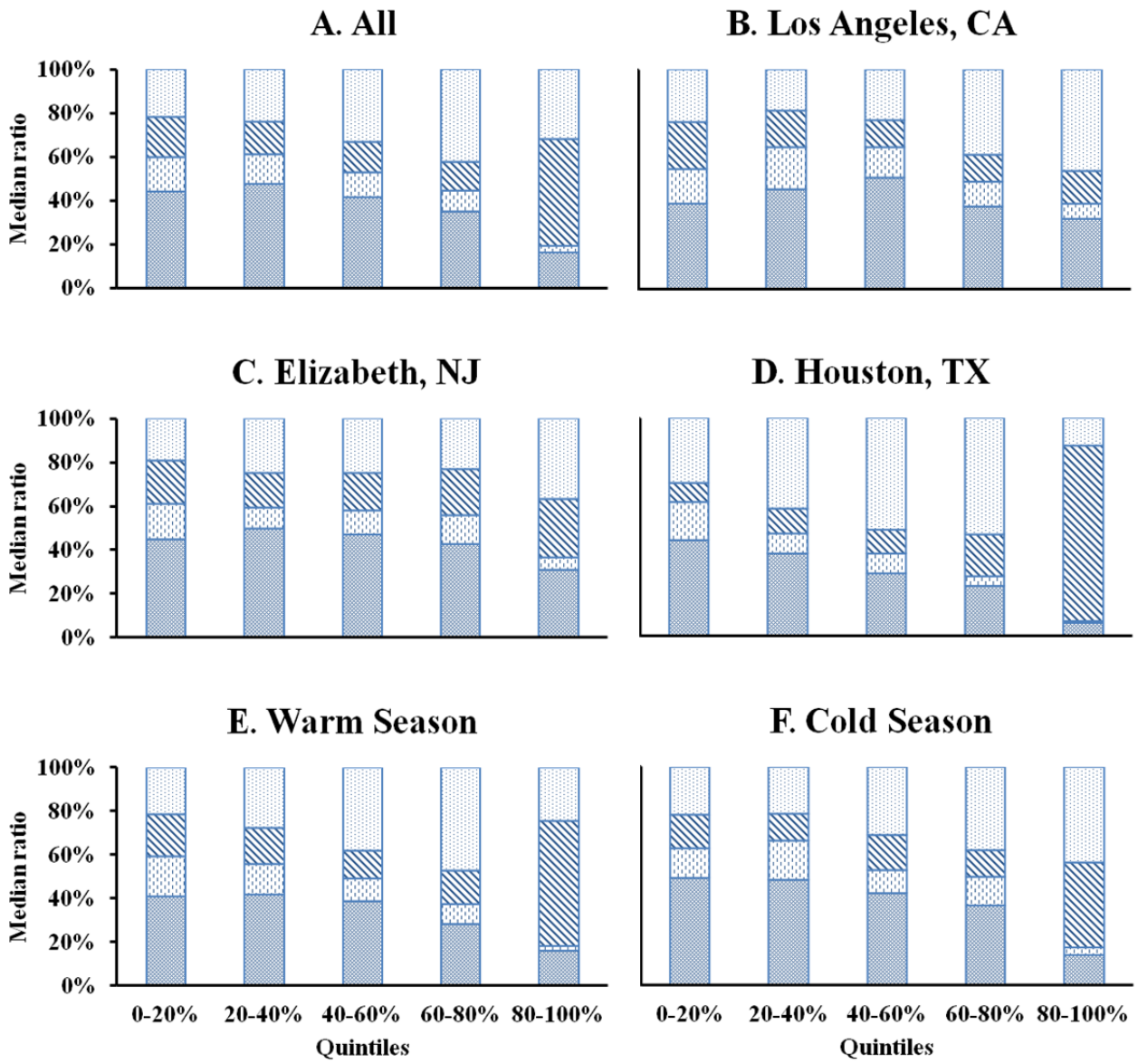

Figure A.11. Indoor VOC composition at quintiles of total VOC concentrations. Warm season indicates April to September, and cold season indicates October to March (n = 554).
, aromatics ; MTBE; chlorocarbons; terpenes.

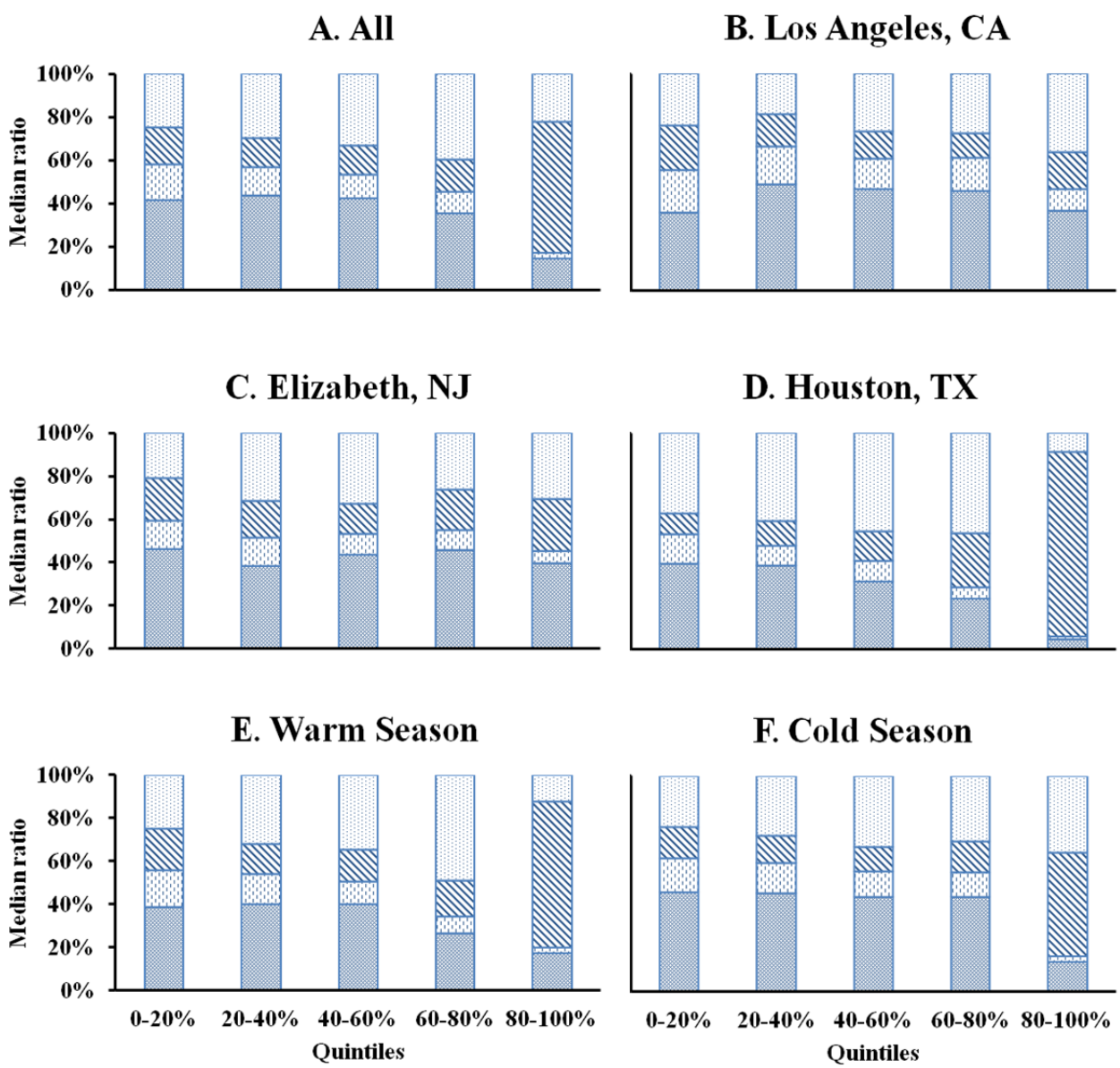

Figure A.12. Personal VOC composition at quintiles of total VOC concentrations. Warm season indicates April to September, and cold season indicates October to March (n = 544).
aromatics ; , MTBE; , chlorocarbons; , terpenes.

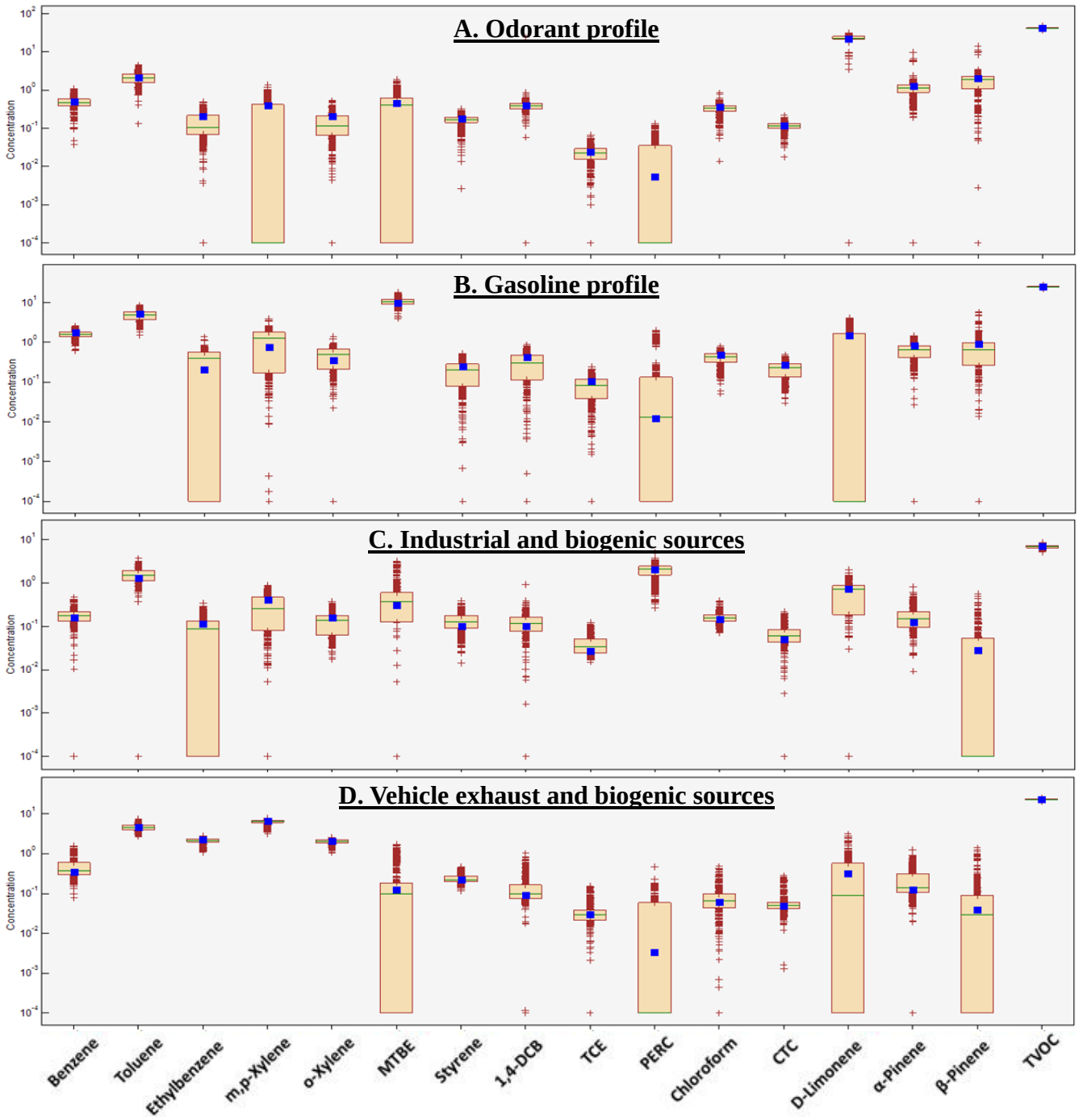

Figure A.13. Factor profiles and variability for personal VOC exposures based on bootstrap analyses using PMF (n = 299). Blue boxes show original factor profiles; bars with red outlines show interquartile ranges; green lines are the medians of the bootstrap results; and red crosses are values outside the interquartile range.

4. Identify High-Exposure Mixtures

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-concentration groups, using a cut-off of the 75th percentile of the mixture's total concentration (sum of all components). 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 AER. The logistic regression models used proc logistic in SAS 9.2 (SAS Institute, Cary, North Carolina, USA).

The analysis of high-exposure mixtures suggested several variables associated with high exposures (Appendix Table A.11). When comparing the top quartile to the remainder of the data, the following variables were significant (95% confidence interval excluding 1, except as noted):

- City effect: Participants in Los Angeles and Elizabeth had lower odds of high exposure (≥ 75th percentile) than Houston participants for all mixtures (odds ratios [ORs] from 0.18 to 0.63), except mixture A3 for the Elizabeth participants.
- Race/ethnicity: Participants of Mexican descent 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 with white subjects (ORs from 2.03 to 3.97). Hispanic subjects had higher odds of high exposure to mixture A3 than white subjects (OR = 1.78, 95% CI = 1.09–2.92). Asian, black, and Indian participants were less likely to have high exposure to mixture A2 (toluene, ethylbenzene, xylene, and styrene) than white subjects (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 , such as *d*-limonene and the pinenes in mixture A4 (ORs from 0.38 to 0.69).
- Open doors or windows: Participants who reported opening doors or windows during the sampling periods had lower odds of high exposure for all mixtures than individuals who did not (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 the pinenes).
- Attached garages: Participants living in houses with attached garages had increased odds of high exposure to mixtures A1 (gasoline vapor) and A2 (vehicle exhaust) (ORs = 2.27 and 1.95, 95% CIs = 1.45–3.56 and 1.25–3.05, respectively).

- Participants' activities: Participants who pumped gas during the sampling period had increased odds of high exposure to the gasoline mixture A1 (OR = 2.10, 95% CI = 1.25–3.52). Participants who used fresheners had higher odds of having high exposure to mixture A4 with *d*-limonene, α-pinene, and β-pinene (OR = 2.20, 95% CI = 1.17–4.14).
- Activities of family members: Participants with family members who showered during the sampling period had increased odds of high exposures to mixture A3 with moth repellents, chlorinated solvents, and water disinfection by-products (OR = 2.06, 95% CI = 1.20–3.56) and to mixture 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. Several factors identified for gasoline and vehicle exhaust mixtures for the RIOPA participants have also been shown in personal exposure measurements in NHANES; for example, the presence of attached garages and pumping gas were related to benzene, toluene, and MTBE exposures (Jia et al. 2010). However, statistically significant factors have not been previously 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 for more time (recruitment requirements for the RIOPA study; 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 for the preceding analysis do not require normality of the response variables. Thus, variables with right-skewed distributions, such as VOC concentrations, 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 (i.e., the determinants of exposure), which accounts for repeated measures and interactions (using linear mixed-effects models), is provided in the main report.

Table A.11. Results of bivariate logistic regression models for VOC mixtures in RIOPA identified by PMF.

Potential factor		Mixtures (n = 299)							
		A1: Benzene and MTBE		A2: Toluene, ethylbenzene, xylenes, and styrene		A3: 1,4-DCB, TCE, PERC, chloroform, and CTC		A4: *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	1/hour	**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.

Tables A.12 through A.15 are supplemental materials for the section on copula analyses in the main text.

Table A.12. Distribution type and parameters fitted to individual VOCs (n = 299).

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.

TableA.13. Goodness-of-fit statistics for copulas based on first-visit VOC measurements in RIOPA (n = 299).

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 five AICs or BICs for each mixture was the best-fit copula, and is shown in bold type.

Table A.14. Parameters and correlation matrices of fitted copulas for VOC mixtures (n = 299).

Mixture	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) (Mixtures A3 and B3 had the same VOC components)

	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 = 3)

	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 A.15. Comparison of mixture fractions for mixtures A3/B3 (top set of rows) and B2 (bottom set of rows) for different copula types (n = 299).

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 may not sum to 1.
Dominant mixture fraction shown in bold.

Tables A.16 through A.22 are supplemental materials for the section on VOC determinants in the main text.

Table A.16. Linear mixed-effects model results for outdoor VOCs using multiple 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	1/knot	4.18	0.46	<.0001	0.1	Number of floors		0.10	0.04	0.009	3.2	Inverse wind speed	1/knot	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						
	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									
	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												
						Air conditioning	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							Outdoor temperature	Q1	0.11	0.09	0.242	-28.6
						Using deordorizers or fresheners	No	0.16	0.09	0.087	9.4		Q2	0.36	0.09	<.0001	8.9
Crawl space	No	-0.15	0.08	0.078	-16.5								Q3	0.03	0.09	0.700	2099.0
													Q4	Reference			
Tobacco products smoked in home	No	0.39	0.21	0.062	-40.8												
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															

For dichotomous variables, the reference group is "Yes".

Table A.17. Linear mixed-effects models for indoor VOCs using multiple 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	1/knot	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
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												
	College or above	Reference				Cooking inside or outside	No	0.18	0.08	0.027	-13.6	Logtransformed AER	1/hour	-0.44	0.05	<.0001	-4.4
Professional cleaning	No	0.17	0.10	0.088	-11.7							Spending awake time at	1st floor	-0.39	0.11	0.001	-1.1
													Others	Reference			
Indoor temperature	°C	-0.04	0.01	0.000	12.0	Vacuuming	No	0.21	0.08	0.013	-21.1						
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						
						Professional cleaning	No	-0.14	0.13	0.292	-50.0						
						Logtransformed AER	1/hour	-0.25	0.05	<.0001	-16.6						
						Unemployed	No	0.20	0.11	0.057	-15.7						

For dichotomous variables, the reference group is "Yes".

Table A.18. Linear mixed-effects models for personal VOCs using multiple 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	1/knot	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						
Heating fuel	Electricity	0.10	0.17	0.558	-47.7							Logtransformed AER	1/hour	-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															
Indoor temperature	°C	-0.05	0.01	<.0001	11.8												
Education	Less than HS	0.13	0.12	0.288	-12.8							Unemployed	No	-0.40	0.14	0.005	15.2
	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															
Attached garage	No	-0.18	0.09	0.050	-4.3												
Pumping gas	No	-0.17	0.08	0.044	4.3												

For dichotomous variables, the reference group is "Yes".

Table A.19. Effect sizes for linear mixed-effects models of personal exposure to gasoline-related VOCs (n = 400 to 530 depending on model).

Variable	Group or Unit of Change	Benzene		Toluene		Ethylbenzene		*m*-&*p*-Xylene		*o*-Xylene		MTBE		Styrene	
		Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE
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	-1.14	1.35	-1.25	1.37	-1.23	1.32				
	Mexican			1.21	1.44	1.21	1.44	1.07	1.46	1.12	1.40				
	Hispanic			1.35	1.45	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	3.56°C	**-1.17**	**1.08**												
Inverse wind speed	0.100/knot	**1.52**	**1.11**			**1.37**	**1.14**	**1.33**	**1.15**	**1.29**	**1.13**	**1.80**	**1.18**		
Log-transformed AER	1.09/hour			**-1.39**	**1.12**	**-1.20**	**1.13**	**-1.26**	**1.13**	**-1.17**	**1.12**	**-1.10**	**1.15**		
Number of floors	2	**-1.35**	**1.16**									**-1.48**	**1.26**		
Number of rooms	2	**-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	810 minutes			**-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**		

AER, air exchange rate; HS, high school. Covariates with P values < 0.05 are shown in bold type.
For continuous variables, the effect size is equal to the change in exposure for one interquartile range of the determinant. For dichotomous variables, the reference group is "Yes".

Table A.20. Effect sizes for linear mixed-effects models of personal exposure to odorant-related VOCs (n = 393 to 433 depending on model).

Variable	Group or Unit of Change	1,4-DCB		Chloroform		*d*-Limonene		α-Pinene		β-Pinene	
		Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE	Effect Size	SE
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	12.7%			**-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	1.09/hour			**-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	2	**-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**				

AER, air exchange rate; CHS, central heating system. Covariates with P values < 0.05 are shown in bold type.
For continuous variables, the effect size is equal to the change in exposure for one inter-quartile range of the determinant. For dichotomous variables, the reference group is "Yes".

Table A.21. Effect sizes for linear mixed-effects models of personal exposure to dry-cleaning and industrial-related VOCs (n = 400 to 446 depending on models).

Variable	Group or Unit of Change	TCE		PERC		CTC	
		Effect Size	SE	Effect Size	SE	Effect Size	SE
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	12.7%			**-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	3.56°C	**-1.10**	**1.10**			1.04	1.04
Inverse wind speed	0.100/knot			**1.63**	**1.18**		
Log-transformed AER	1.09/hour			**-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	120 minites	**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**		

AER, air exchange rate. Covariates with P values < 0.05 are shown in bold type.

For continuous variables, the effect size is equal to the change in exposure for one inter-quartile range of the determinant. For dichotomous variables, the reference group is "Yes".

Table A.22. Reduction in the residual variance (R^2) attributable to fixed-effect variables in the linear mixed-effects models for the RIOPA VOCs.

VOCs	R^2		
	Outdoor (n = 555)	Indoor (n = 554)	Personal (n = 544)
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

Figures A.14 and A.15 are supplemental information for the section on determinant analyses in the main text.

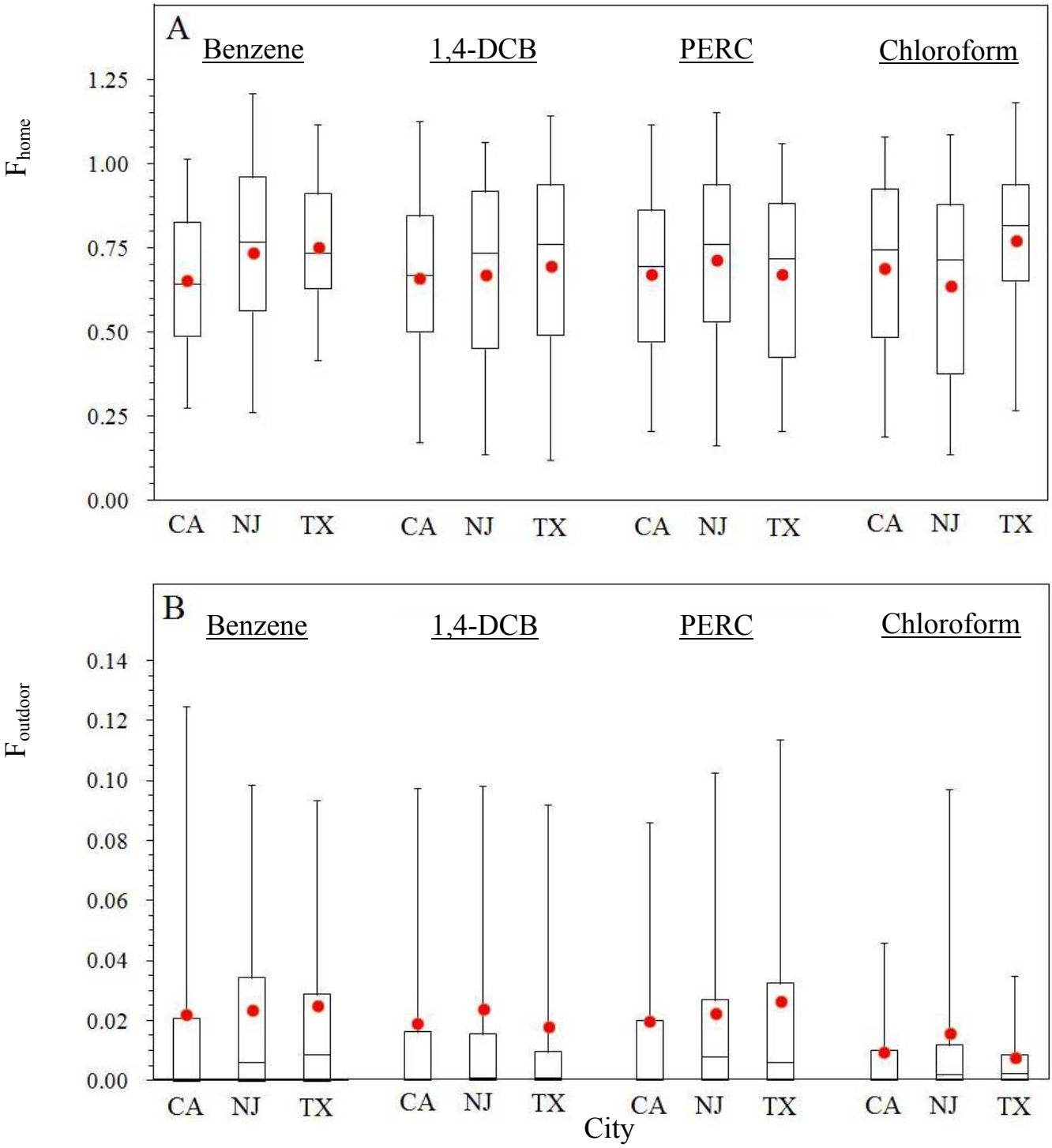

Figure A.14. Box plots showing F_{home} and $F_{outdoor}$ for selected VOCs in the three RIOPA cities (CA, Los Angeles; NJ, Elizabeth; TX, Houston). Plots show 5th, 25th, 50th, 75th and 95th percentile concentrations, and average concentration as red dots (n = 544).

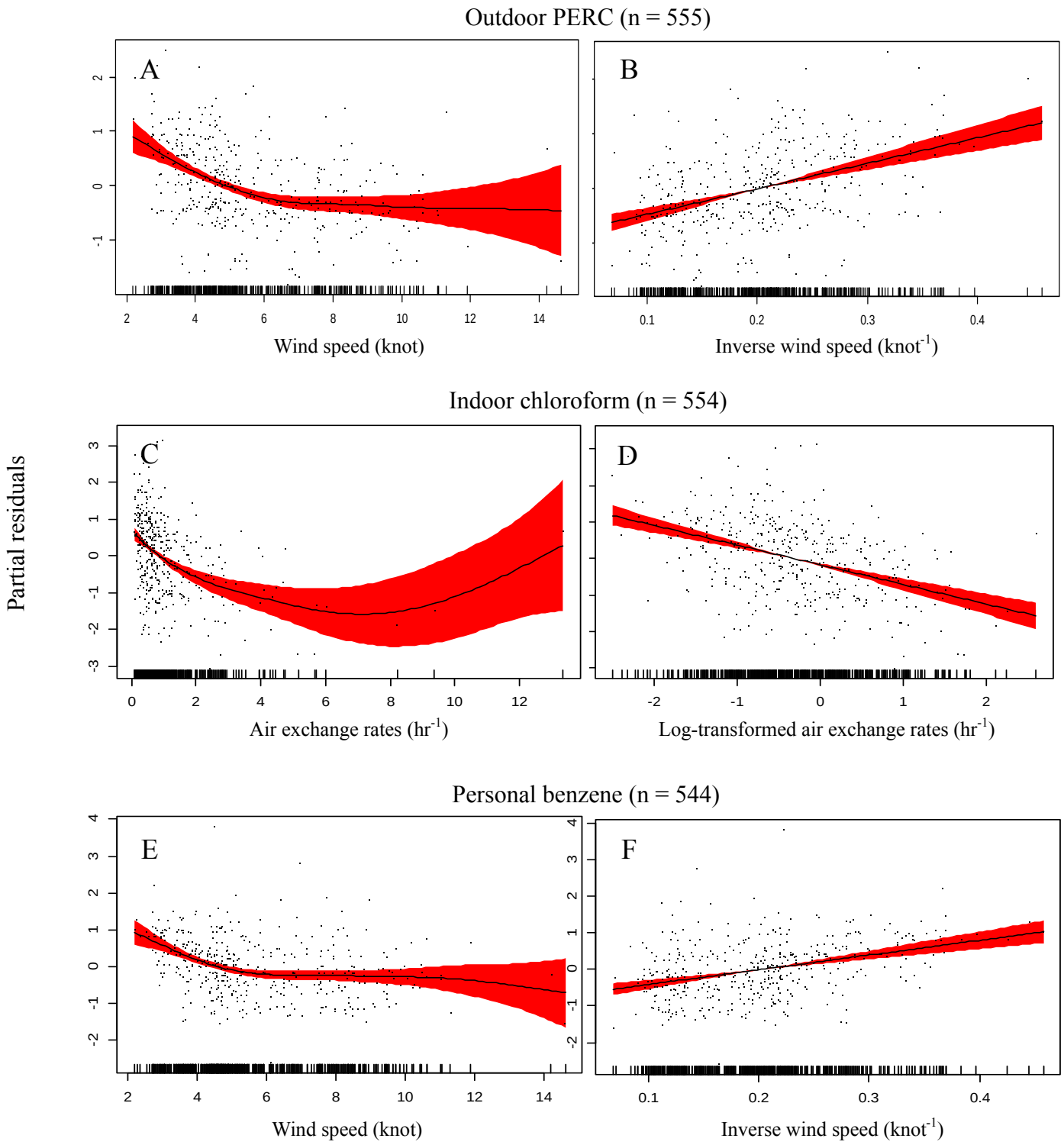

Figure A.15. Partial residual plots of linear mixed-effects models for selected VOCs.

5. References

Agency for Toxic Substances and Disease Registry. 2013. Minimal Risk Levels (MRLs) for Hazardous Substances. Atlanta, GA: Agency for Toxic Substances and Disease Registry. Available: http://www.atsdr.cdc.gov/mrls/mrllist.asp.

Barnett V. 1975. Probability plotting methods and order statistics. Appl Stat 24:95–108.

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

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

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

Delgado-Saborit JM, Aquilina NJ, Meddings C, Baker S, Harrison RM. 2009. Model development and validation of personal exposure to volatile organic compound concentrations. Environ Health Perspect 117(10): 1571–1579.

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

Haas C. 1997. Importance of distributional form in characterizing inputs to Monte Carlo risk assessments. Risk Anal 17:107–13.

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

International Agency for Research on Cancer. 2014. Agents Classified by the IARC Monographs. Lyon, France: World Health Organization, IARC. Available: http://monographs.iarc.fr/ENG/Classification/index.php.

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

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

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

Office of Environmental Health Hazard Assessment. 2005. Air Toxics Hot Spots Program Risk Assessment Guidelines, Part II: Technical Support Document for Describing Available Cancer Potency Factors. Sacramento, CA:California Environmental Protection Agency, OEHHA, Air Toxicology and Epidemiology Section.

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

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

Stephens MA. 1974. EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc 69:730–737.

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

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

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

U.S. Environmental Protection Agency. 2001. Trichloroethylene Health Risk Assessment: Synthesis and Characterization (external review draft). National Center for Environmental Assessment, Office of Research and Development.

U.S. Environmental Protection Agency. 2008. EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. Available from *www.epa.gov/heasd/products/pmf/EPA%20PMF%2030%20User%20Guide%20v16_092208_finalpdf.*

U.S. Environmental Protection Agency. 2010. Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene. Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, The National Academies Press.

U.S. Environmental Protection Agency. 2014. Integrated Risk Information System (IRIS). Washington, DC: US Environmental Protection Agency Available: http://wwwepagov/iris/search_humanhtm.

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

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

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