Breath Biomarkers in Environmental Health Science: Exploring Patterns in the Human Exposome

Joachim D. Pleil[a]*, Matthew A. Stiegel[b], and Jon R. Sobus[a]

[a]Human Exposure and Atmospheric Sciences Division, NERL/ORD, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711;

[b]Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, 27599

Abstract

The human genome is the counterpart to the human exposome with respect to the gene x environment interaction that describes health state and outcome. The genome has already been sequenced and is in the process of being assessed for specific functionality; to similarly decode the exposome will require the measurement and interpretation of suites of biomarker compounds in biological media such as blood, breath, and urine. Of these, exhaled breath provides some important advantages for community or population based studies in that the supply is essentially unlimited, the sampling procedures are non-invasive and can be self-administered, and there are little, if any, infectious wastes generated. The main concerns are to document a variety of compounds in breath, to assess what compounds and concentrations are considered statistically "normal" in the healthy or unremarkably exposed population, and what graphic and mathematical approaches can be applied to assess outlier measurements as perturbations to the healthy exposome. In this paper, we explore a data set of exhaled breath measurements of exogenous exposures to jet fuel and develop summary statistics and variable clustering methods to establish between-group and intrinsic within-sample patterns that could be used to assess the status of random subjects.

Introduction

Conventional wisdom proposes that environmental exposures contribute to adverse human health outcomes; however, with few exceptions (cigarettes, asbestos, lead, radon, benzene, etc.), it is difficult to unambiguously demonstrate the direct linkage between exposure and a specific disease (Pleil and Sheldon 2010). As such, the general concepts regarding human health are that the risks of developing chronic diseases are attributable to an unspecified mixture of genetic and environmental factors. Discussions in the literature regarding recent advances in disease etiology and epidemiology have suggested that cardiovascular disease, autoimmune diseases, cancer, and other apparently random long-term latency diseases may have causal links among the genome, metabolism, life-stage, and the environment (Rioux and Abbas 2005, Rappaport 2010, Pleil et al. 2011a, Lahiri 2011, Wild 2011). Furthermore, Rappaport and Smith (2011) have recently asserted that *"…70 to 90% of disease risks are probably due to differences in environments."* We note that in this definition, they define "environment" as anything that is not the genome, and thus it includes the traditional ingestion, inhalation, and dermal exposures, as well as internal chemistry from reactive oxygen species, phase-1 metabolism, signaling molecules, and all forms of waste products from oxygen based energy production. This complementary side to the genome has been termed the "exposome" and, according to Wild (2005), *"…at its most complete, the exposome encompasses life-course environmental exposures (including lifestyle factors), from the prenatal period onwards."*

The human exposome, therefore, is a window not only into the "gene x environment interaction" (Smith and Rappaport 2009), but also serves as an indicator of previous environmental exposure and current health state. Certainly, one cannot measure every chemical (exogenous and endogenous) in the human body, but we have begun to piece together suites of biomarker compounds that can be interpreted as partial snapshots of the exposome. For example, we have accumulated summary statistics (e.g. means, medians, distributions, internal concentrations, elimination kinetics) for polycyclic aromatic hydrocarbons (PAHs) in blood (Pleil et al. 2010), PAHs and hydroxyl-PAHs metabolites in urine (Sobus et al. 2010), volatile alcohols and aldehydes in exhaled breath condensate (EBC) (Pleil et al. 2008, Hubbard et al. 2009), alkanes and single-ring aromatics in exhaled breath (Pleil 2009, Pleil et al. 2011b), C_2 to C_{12} carbonyls and total protein in EBC and bronchial lavage fluid (BLF) (Sawyer et al. 2008), and we are currently working on a new series of projects measuring inflammatory cytokines in human plasma, urine, and EBC.

As discussed above, gas-phase exhaled breath and EBC have become an integral part of exposome investigations and documentation. Although blood and urine are the traditional biological media for systems biology experiments, breath media sampling is generally preferred in environmental health science where larger numbers of subjects are studied in contrast to medical or clinical studies. The advantages of breath media revolve around three basic premises: acceptance, as breath sampling is non-invasive; logistics, as sampling does not require medical personnel (blood/BLF) or privacy (urine), and simplicity/safety as breath sampling generates little if any infectious wastes (Lindstrom and Pleil 2002). Furthermore, breath offers a nearly inexhaustible sample supply for time-based analyses, and, unlike blood or urine, can be sampled in rapid succession (essentially in adjacent breaths) to establish uptake and elimination kinetics (Pleil et al. 1998a, Pleil et al. 1999, Lindstrom and Pleil 1996). Finally, breath serves a dual purpose in exposure science investigations: as a surrogate for the circulating blood

concentrations for volatile substances (Pleil et al. 1998b), and also as a "mass balance" tool for uptake and elimination of exogenous compounds and endogenous waste products (Pleil 2008).

Biomarker Framework

There are hundreds to thousands of discrete chemicals in any given exhaled breath sample that serve as indicators of the internal human systems biology (Pauling et al. 1971, Phillips et al. 1999, Cao and Duan 2006). We have recently developed a framework document for environmental biomonitoring research that illuminates how measurements fit into the sequence from environmental stressor to internal dose to eventual health outcome (Sobus et al. 2011). As part of the framework, empirical measurements (boxes) and estimated values (triangles) exist along parallel tracks, and are connected by various modeling linkages (see Figure 1). The estimated values, which include "exposure" and "dose", are the most commonly used metrics for regulatory risk assessment. Yet, measurement data from external and internal environments are always needed to generate and confirm these values. As such, breath data can provide important information along the measurement progression. For example, in the box labeled "Biomarker measurements", we can insert the chemical biomarkers such as native compounds of exposure, which directly link back to the environmental exposure. In the box labeled "BR Biomarker measurements" (where BR refers to biologically relevant), we can insert measurements of phase-1 metabolites of environmental contaminants, or endogenous chemicals such as aldehydes, ketones, alcohols, thiols, hydrocarbons, amines, etc., that may reflect biologically active dose or pre-clinical health effects.

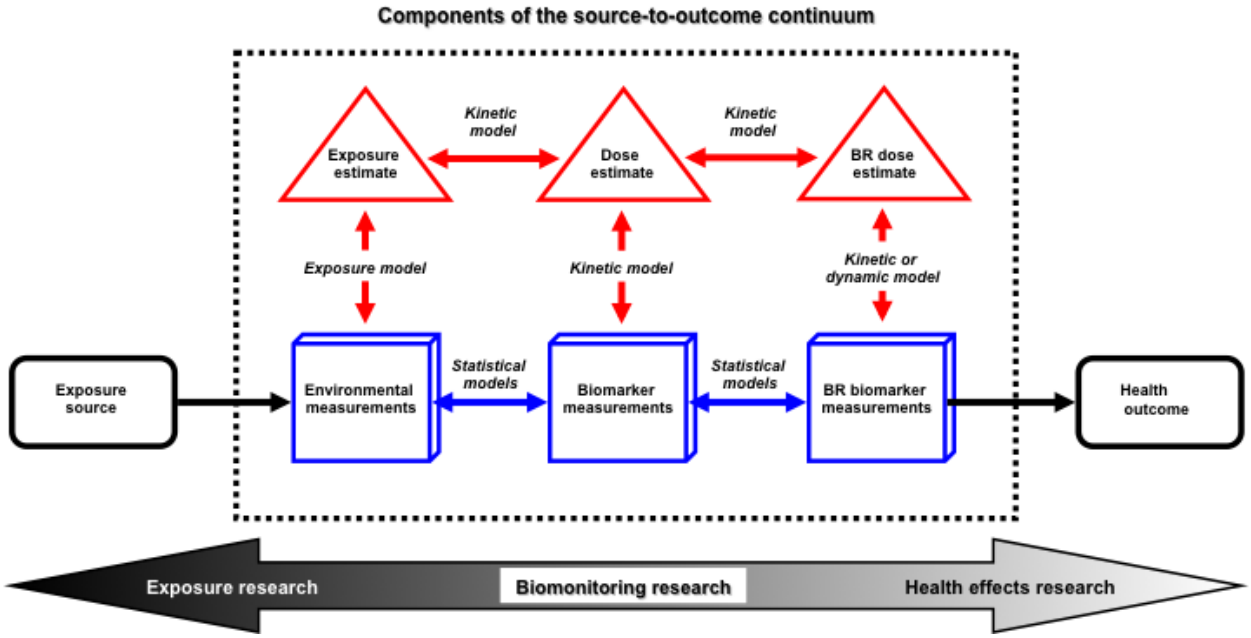


Figure 1: Framework diagram of the exposure source to health outcome continuum showing parallel tracks of measurements (boxes), estimated/calculated values (triangles) and the respective modeling linkages (arrows). Breath measurements fit into the boxes labeled "biomarker measurements" and "BR biomarker measurements". The center box contains measurements of exhaled compounds of environmental exposure whereas the "BR" box contains measurements of biologically relevant exhaled compounds indicating metabolism, perturbations, and pre-clinical effects.

Biomarker interpretation

As discussed above, breath biomarkers are most valuable for these systemic approaches in a pattern format (i.e. 10 or more compounds measured at a time). We have found that focusing on series of "*one-compound exposure – one-compound metabolite biomarker*" models is no longer completely sufficient as the human biological system is generally interacting with thousands of exogenous and endogenous chemicals simultaneously. In the new exposome context, we consider sets of measured compounds as groups, and then interpret them using various pattern recognition schemes including graphical/visualization techniques (e.g. network diagrams and heat maps), and computational techniques including principal components analysis (PCA), variable clustering (VarClus), general linear models (GLM), and mixed-effects regression models. These new approaches all require great care in assigning probative value to their outcomes.

One of the most difficult challenges in quantifying patterns is the fundamental statistical problem of over-modeling, that is, when the ratio of samples to putatively independent variables is less than about 10, it becomes difficult to determine underlying distributions and to get unambiguous results. The obvious solution is to collect more samples; however, this is not often possible from a resources or subject availability perspective. Often, we look at the patterns

qualitatively without statistical evaluation and try to narrow down the data in some fundamental manner such as: females have higher levels than males; polar compounds are distributed differently than non-polar compounds, etc. We also look at within-sample covariance wherein we assess if compound subsets co-vary or if compounds are truly independent.

Biomarker data

As an example, consider the following exhaled breath data of jet-fuel exposures extracted from a previous publication (Pleil et al. 2011b). Table 1 shows the exhaled breath summary data organized into 4-groups: pre-shift and post shift control subjects (n=37 each); pre-shift and post-shift exposed subjects (n=114 each). Work shift is defined as 4-hours performing normal assigned tasks, and exposure status is determined by whether or not the subject's normal task required hands-on jet fuel systems maintenance. By inspection, we see that pre-shift levels are lower than post shift levels, and that subjects working directly with fuel generally have higher internal doses. Performing any meaningful statistics, however, is fraught with difficulty as there are insufficient numbers of samples to determine a model based on 12 compounds with only 37 or 114 samples per group, respectively, across a total of 7-bases and one-week (5 working days) snapshots across a 6-month overall period. However, we expect that all pre-shift samples, plus the post-shift controls, should reflect the generic Air Force Base (unremarkable) population. One-way ANOVA analysis of the log-space results shown in Table 1 confirms this conjecture; ($p = 0.255$ for means and $p = 0.716$ for variances with Bonferroni correction for repeat analyses) and so we can use these three groups together as a basis set. An additional analysis that included the post-shift exposed cohort demonstrated a highly significant difference for means ($p < 0.0001$) but the same variance statistics ($p = 0.871$).

Table 1: Summary statistics for exhaled breath concentrations (ppbv) and 95% confidence intervals for different sample groups.

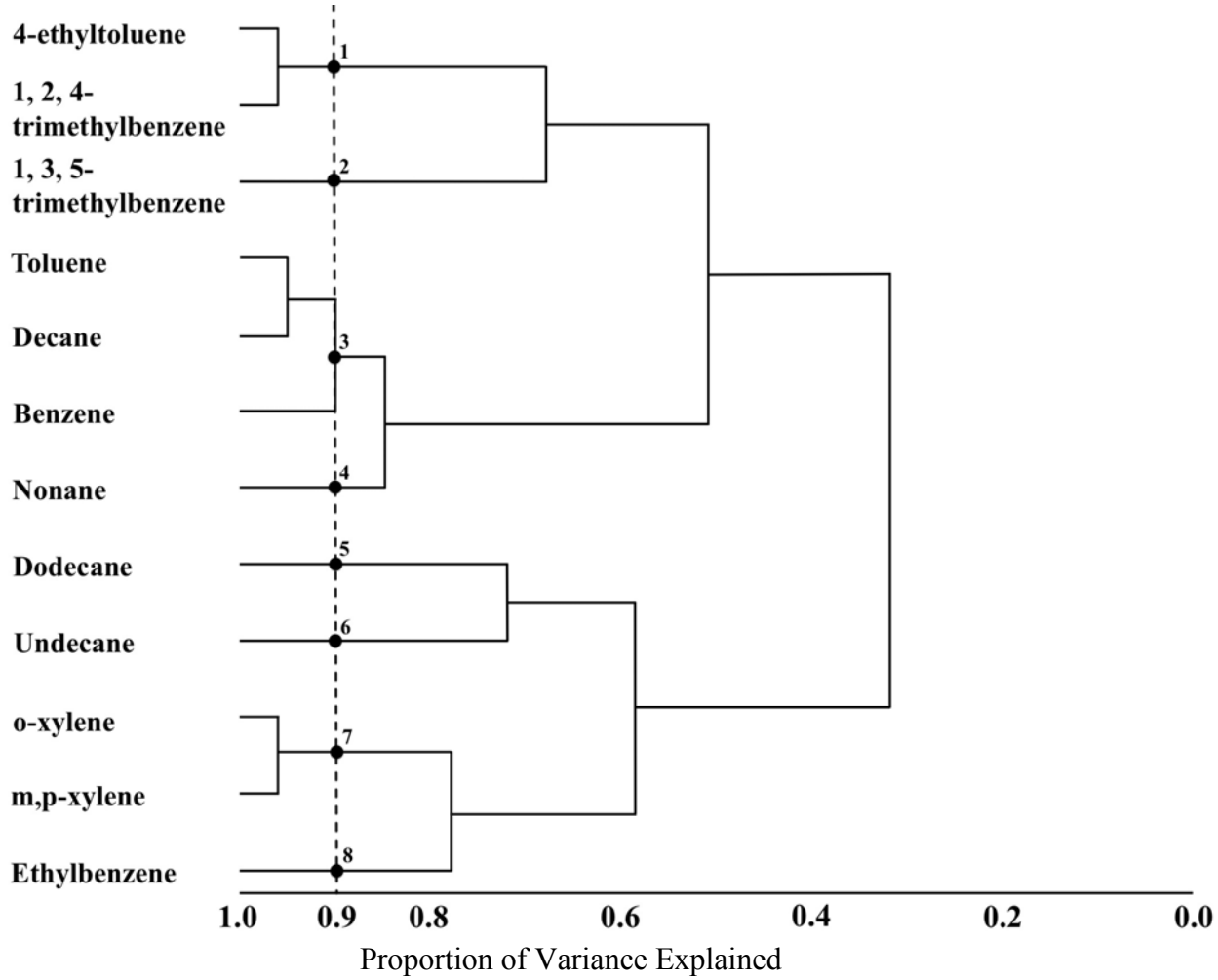
	Pre-shift controls	Post-shift controls	Pre-shift exposeds	Post-shift exposeds
Compounds	median ppbv n=37 (95% CI)	median ppbv n=37 (95% CI)	median ppbv n=114 (95% CI)	median ppbv n=114 (95% CI)
benzene	0.55 (0.08, 3.3)	0.51 (0.07, 3.8)	0.73 (0.10, 5.1)	2.22 (0.18, 27)
ethylbenzene	0.19 (0.03, 1.3)	0.18 (0.03, 1.3)	0.25 (0.05, 1.4)	3.28 (0.37, 29)
1,3,5-trimethylbenzene	0.09 (0.01, 1.0)	0.08 (0.01, 1.3)	0.19 (0.02, 1.6)	2.74 (0.21, 36)
o-xylene	0.18 (0.02, 1.4)	0.16 (0.02, 1.4)	0.30 (0.5, 1.7)	4.86 (0.35, 68)
4-ethyltoluene	0.17 (0.03, 1,2)	0.23 (0.02, 2.7)	0.35 (0.04, 28)	6.43 (0.56, 73)
1,2,4-trimethylbenzene	0.29 (0.04, 2.1)	0.36 (0.03, 4.8)	0.54 (0.07, 4.1)	7.95 (0.67, 94)
m,p-xylene	0.51 (0.06, 3.6)	0.46 (0.06, 3.6)	0.71 (0.13, 3.8)	8.67 (0.80, 94)
toluene	2.05 (0.55, 7.60)	1.96 (0.37, 10)	1.96 (0.23, 17)	8.43 (0.69, 103)
dodecane	0.42 (0.02, 9.3)	1.26 (0.07, 22)	0.93 (0.07, 12)	13.06 (1.7, 100)
undecane	0.75 (0.04, 13)	1.53 (0.12, 19)	1.62 (0.23, 11)	19.49 (2.29, 166)
decane	0.55 (0.04, 8.0)	0.90 (0.05, 16)	1.40 (0.18, 11)	24.78 (2.6, 240)
nonane	0.33 (0.01, 7.6)	0.50 (0.02, 15)	0.97 (0.08, 12)	25.87 (2.3, 297)

Biomarker clustering

The covariance of individual compounds can be explored using a variable clustering (eigenvector) approach which allows us to combine compounds that behave similarly across samples (Domany 2003, Kettenring 2006, Pleil and Lorber 2007). We performed standard variable cluster analyses (using Proc VarClus, SAS Inc, Cary, NC, USA) for the data groups from Table 1. The results are as expected: all compounds are somewhat correlated within an "r^2" of about 0.4 regardless of grouping. Under the hypothesis (discussed earlier) that all pre-shift samples plus the post-shift controls should represent the unremarkably exposed Air Force

population, and that only the post-shift exposed cohort should have remarkable jet-fuel dose, we combined the former samples (n=188) and contrasted them to the latter exposed samples (n=114). We found overlap in aggregate exposure values between the groups and so excluded samples with >30 ppbv from the first group, and samples with <30 ppbv from the second group resulting in reduced sample sets of n=161 and n=100, respectively. Within these groups, we normalized each sample by the sum of the 12 compounds concentrations to create parameters independent of absolute value. As such, we remove one degree of freedom, but can now compare samples with respect to intrinsic pattern of the compounds. We then used Proc Varclus to assess covariance of the hydrocarbons and aromatics patterns within each group as shown in the dendrite diagrams in Figures 2a and 2b.

A.



B.

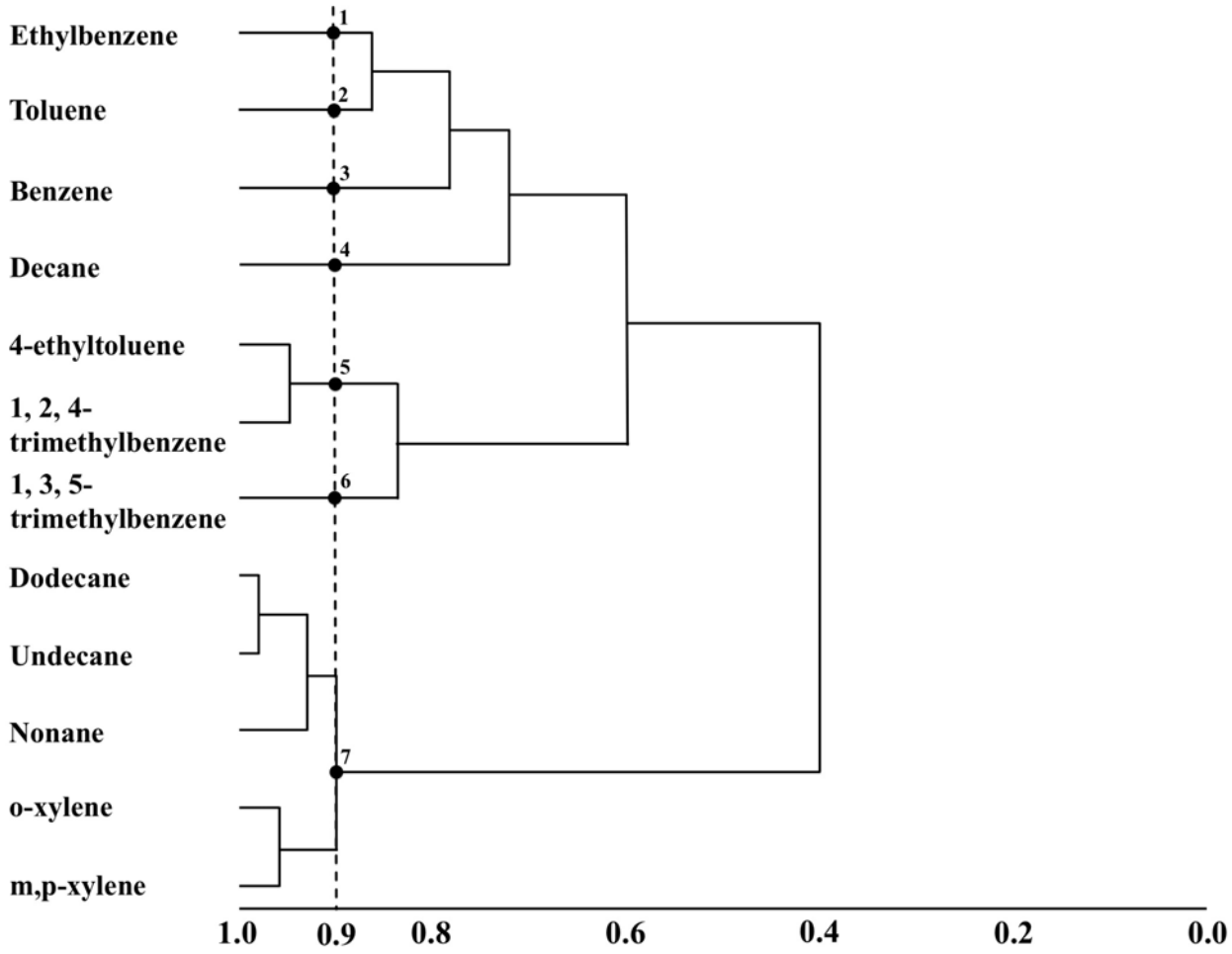


Figure 2. Dendrite diagrams indicating covariance clusters of individual JP-8 constituents. Figure 2a shows clusters for "unremarkable" group and Figure 2b shows clusters for "exposed" group; x-axis indicates proportion of variance explained. In both cases, clusters are chosen to explain a proportion of about 0.9 of the variance (dotted vertical line); in 2a, this results in 8 clusters, in 2b, this results in 7 clusters as indicated by the numbered nodes.

Regression (pattern) models

Once the independent variables are established as clusters, we make use of those imbedded patterns to interpret the exposome and any potential shifts. In general, either of the two Varclus clustering patterns (Figure 2a or 2b) should work equally well to set the baseline pattern; in this case, we will use the "unremarkable" group as the control pattern against which the level of exposure can be evaluated.

The first step is to establish which, if any, of the variable clusters distinguishes between normal baseline and outliers. From a visual comparison of dendrite diagrams in Figures 2a and 2b, we note that the patterns are indeed different, and so we expect that there should be a mathematical method for extracting an algorithm to separate the two groups by pattern. To establish a parsimonious regression equation without losing too much explanatory power, we choose to form eight clusters (C_i), which, in aggregate, explain a proportion of about 0.9 (or 90%) of the variance (see dotted vertical line in Figure 2a). Although 90% or 95% are commonly chosen discriminators for model development, this is somewhat arbitrary because it is a trade-off that often depends on the data structure and the number of available samples. There are times when 80% or lower may be the best choice.

For our example here, we have 90% explained variance using 161 samples and 8 clusters. This results in a "comfortable" sample/variable ratio of ~ 20 which is well above the rule of thumb limit of 10 (Pleil and Lorber 2007). Using the first group (Figure 2a), we establish the following eight clusters, C_i, at 90% explained variance:

Group 1 clusters: "unremarkable"

C_1 = 4-ethyltoluene + 1,2,4- trimethylbenzene
C_2 = 1,3,5- trimethylbenzene
C_3 = toluene + decane + benzene
C_4 = nonane
C_5 = dodecane
C_6 = undecane
C_7 = o-xylene + m,p-xylene
C_8 = ethylbenzene

Similarly, we could have established clusters for the second group that also explain ~90% of the variance (Figure 2b) as follows:

Group 2 clusters: "exposed"

C_1 = ethylbenzene
C_2 = toluene
C_3 = benzene
C_4 = decane
C_5 = 4-ethyltoluene + 1,2,4-trimethylbenzene
C_6 = 1,3,5-trimethylbenzene
C_7 = dodecane +undecane + nonane + o-xylene + m,p-xylene

As mentioned above, we choose to use the clusters from Group 1 as the baseline because, in the long term, we want to assess changes from normal or "unremarkable" exposures.

The second step is to establish a parsimonious regression equation wherein only the important clusters are used as independent variables so that we can estimate appropriate regression coefficients. We employ multivariate regression analysis of the form:

$$OPT(S_i) = a_0 + \Sigma(a_n \times C_{n,i})$$

wherein OPT is the output variable for the "i^{th}" sample (S_i), a_0 is the intercept, a_n is the n^{th} cluster regression coefficient, and $C_{n,i}$ is the sum of the analyte concentrations represented by the n^{th} cluster evaluated for the i^{th} sample. Note that for determining the coefficients, we set $OPT(S_i) = 0$ for the "unremarkable" group and $OPT(S_i) = 1$, for the exposed group. At $\alpha = 0.05$, only clusters C_3, C_5, C_6 and C_7 survive, and the regression equation results in:

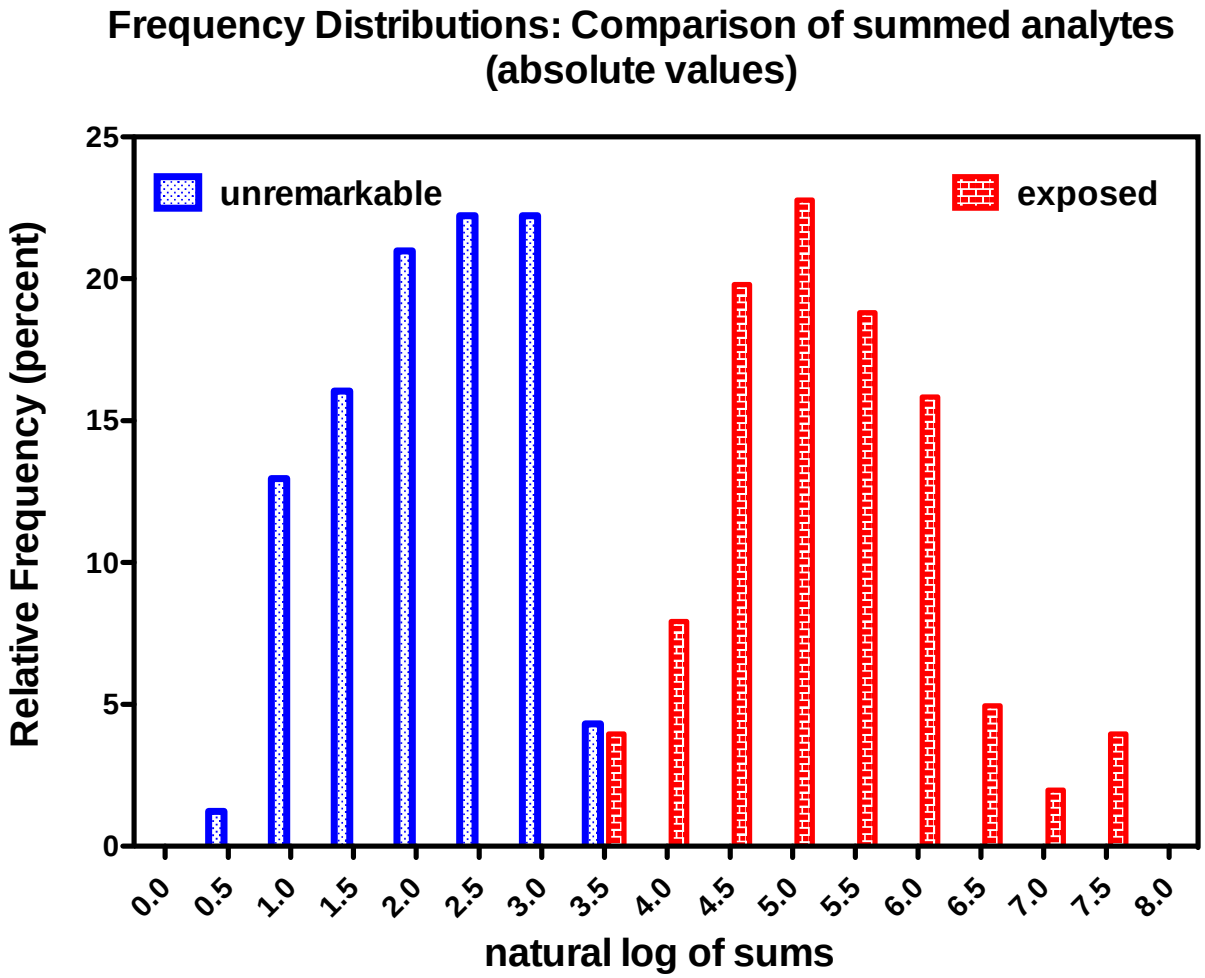
$$OPT(S_i) = 2.95 - (3.32 \times C_{3i}) - (2.31 \times C_{5i}) - (3.36 \times C_{6i}) - (5.49 \times C_{7i})$$

Once this equation is established for a set of basis data (in this case for "unremarkable" and "exposed" sample groups), OPT(S) is treated as a continuous variable for an unknown sample "S_u" and the calculated result can be interpreted to see which of the two extremes S_u most resembles based on the intrinsic pattern. We would expect that calculated results of $OPT(S_u)$ for unknown samples will be closer to 0 for unremarkably exposed subjects and closer to 1 for exposed subjects.

Biomarker group comparisons

There are two different scenarios for visualizing the "unremarkable" and the "exposure" groups; the first shown in Figure 3a is by absolute concentrations of the compounds as shown in Table 1 (in this case we look at their sums); the second is shown in Figure 3b by their contribution clusters (indicator of intrinsic pattern) as established by multivariate regression calculations. We see that in the case of subjects that work at Air Force Bases, the exogenous jet fuel biomarkers are somewhat better resolved by absolute amount rather than by intrinsic pattern. This is expected as there is always some low level exposure regardless of specific activity and that those who handle fuels directly will have a higher body burden. What is interesting is that the intrinsic pattern (after absolute value is normalized) still shows moderate separation of the groups indicating that there are sources for other petrochemicals (e.g. gasoline, diesel, lubricating/heating oil) beyond the jet fuel source.

A.


Frequency Distributions: Comparison of summed analytes (absolute values)
unremarkable
exposed
Relative Frequency (percent)
25
20
15
10
5
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
natural log of sums

B.

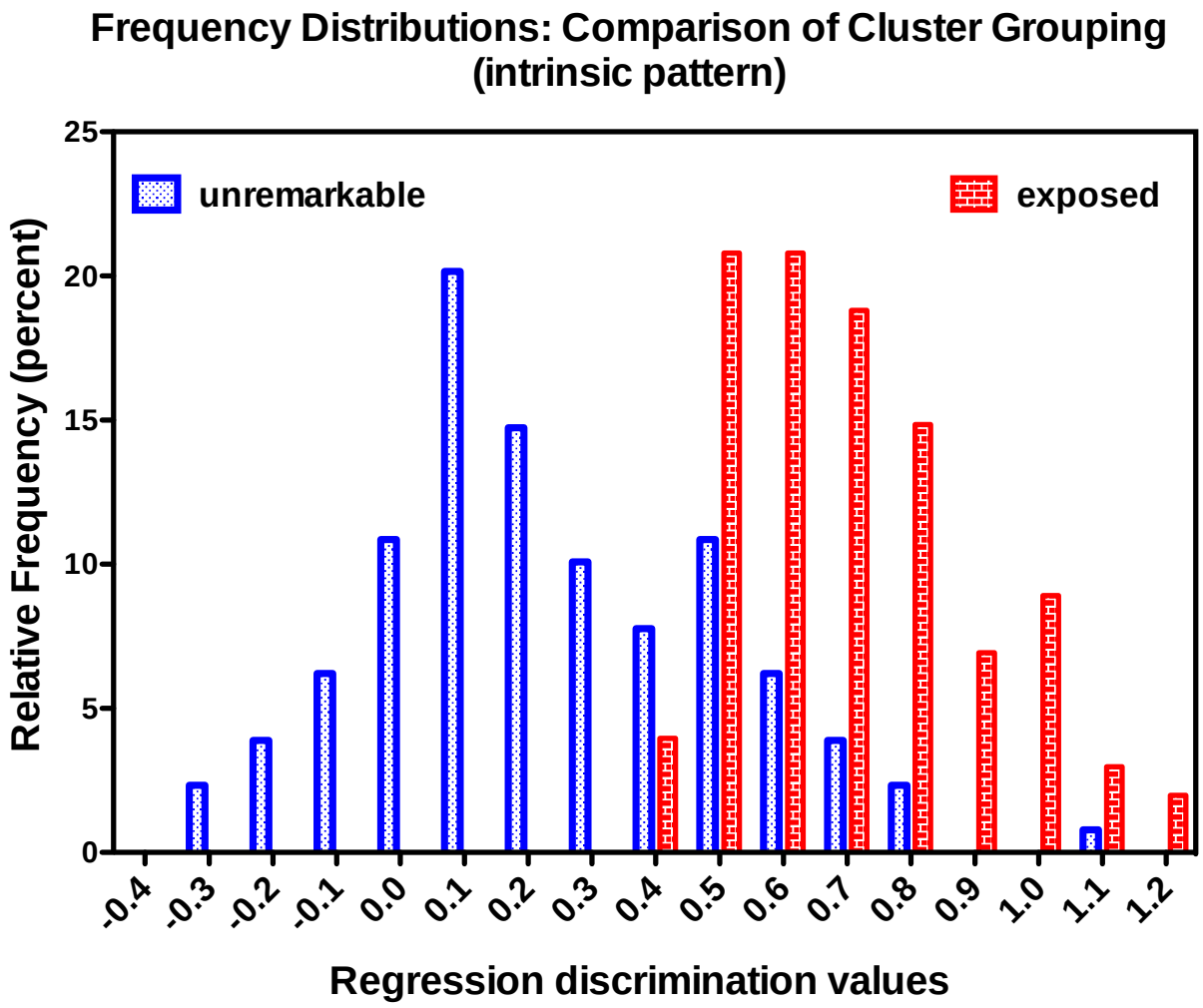


Figure 3. Frequency distribution comparisons for "unremarkable" and "exposed" sample groups. Figure 3a shows separation of groups by absolute concentrations of jet fuel constituent sums and Figure 3b shows separation of groups using intrinsic patterns after removing the absolute concentration differences by normalization.

Certainly the two basis groups (unremarkable and exposed) in Figure 3a are separated by the absolute value of measured hydrocarbons; they were defined as such. However, once the absolute value is normalized away, the combination of variable clustering technique and multivariate regression modeling leaves behind only the intrinsic (internal) pattern in Figure 3b. This is an important feature for pattern recognition as the absolute level of specific compounds can be affected by parameters that are not necessarily related to a specific source, but more by the proximity to a similar source.

Conclusions

Patterns in the human exposome are elusive and their interpretation requires both visual and mathematical approaches. Under the basic exposure science framework construct shown in Figure 1, we have shown how empirical measurements of chemicals representing parts of the human exposome fit into the overall exposure to effect progression in parallel to statistical and kinetic modeling. We demonstrated that absolute value patterns of compounds (as shown in Table 1) and intrinsic patterns using cluster analysis and multivariate regression (as shown in Figures 2 and 3), can be used to assess group differences among biomarker samples. The jet fuel breath experiment presented here is only one example; these approaches can be employed anytime we have groups of complex exposome biomarker data. For example, biomarker data from exploratory case-control experiments of cancer patients could be assessed using cluster analysis and regression to develop better diagnostic procedures. Another potential application would be to test results from different "healthy" control groups thought to be nominally identical to assess if they are actually representing subtly different patterns. For example, are measurements from general population control subjects similar to control subjects drawn from the unaffected hospital population?

The primary value of using these methods is that we address two important problems in complex data analysis: over-modeling (i.e. drawing conclusions based on random effects) and co-linearity among putatively independent variables (resulting in ambiguous conclusions). The ultimate conclusion from this work is that it is possible to take complex data sets and establish a fairly simple comparison metric to assess where unknown samples fall in a continuum from "unremarkable" to "remarkable" using measurements of subsets of chemicals within the human exposome.

Acknowledgements

The authors wish to thank Tzipporah Kormos, Andrew Lindstrom, and Myriam Medina-Vera of US EPA for their expert advice. This research has been subjected to (EPA) Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Cao W, Duan Y, 2006. "Breath analysis: Potential for clinical diagnosis and exposure assessment", Clinical Chemistry 52(5):800-811.

Domany E. 2003. "Cluster analysis of gene expression data", Journal of Statistical Physics 110(3-6):1117-1139.

Hubbard HF, Sobus JR, Pleil JD, Madden MC, and Tabucchi S, 2009. "Application of novel method to measure endogenous VOCs in exhaled breath condensate before and after exposure to diesel exhaust", Journal of Chromatography, B 877(29):3652-3658.

Kettenring JR, 2006. "The practice of cluster analysis", Journal of Classification, 23:3-30.

Lahiri DK, 2011. "An integrated approach to genome studies", Science 331(6014):47.

Lindstrom AB, Pleil JD, 1996. "A methodological approach for exposure assessment studies in residences using volatile organic compound contaminated water", JAWMA 46(11):1058-1066.

Lindstrom, AB and Pleil JD, 2002. "A review of the USEPA's single breath canister (SBC) method for exhaled volatile organic biomarkers", Biomarkers, 7(3):189-208.

Pauling L, Robinson AB, Teranishi R, and Cary P, 1971. "Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography", Proceedings of the National Academy of Sciences USA, 68(10):2374-2376.

Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, and Cataneo RN, 1999. "Variation in volatile organic compounds in the breath of normal humans", Journal of Chromatography, B 877(29):3652-3658.

Pleil JD and Lindstrom AB, 1997. "Exhaled human breath measurement for assessing exposure to halogenated volatile organic compounds", Clinical Chemistry, 43:5 723-730.

Pleil JD and Lindstrom AB, 1998a. "Sample timing and mathematical considerations for modeling breath elimination of volatile organic compounds", Risk Analysis, 18(5):585-602.

Pleil JD, Fisher JW, and Lindstrom AB, 1998b. "Comparison of human blood and breath levels of trichloroethylene from controlled inhalation exposure", Environmental Health Perspectives 106(9): 573-580.

Pleil JD and Lorber MN, 2007. "Relative Congener Scaling of Polychlorinated Dibenzo-p-dioxins and Dibenzofurans to Estimate Building Fire Contributions in Air, Surface Wipes, and Dust Samples", Environmental Science and Technology, 41(21):7286-7293.

Pleil JD, Hubbard HF, Sobus JR, and Madden MC. 2008. "Volatile polar metabolites in exhaled breath condensate (EBC): collection and analysis", Journal of Breath Research, 2: 026001 (9pp).

Pleil JD, 2008. "The Role of Exhaled Breath Biomarkers in Environmental Health Science", Journal of Toxicology and Environmental Health: Part B: Critical Reviews, 11:613-626.

Pleil JD, 2009. "Influence of systems biology response and environmental exposure level on between-subject variability in breath and blood biomarkers", Biomarkers 14(8):560-571.

Pleil JD and Sheldon LS, 2011. Adapting Concepts from Systems Biology to Develop Systems Exposure Event Networks for Exposure Science Research, Biomarkers 16(2):99-105.

Pleil JD, Stiegel MA, Sobus JR, Tabucchi S, Ghio AJ, and Madden MC, 2010. **"**Cumulative exposure assessment for trace-level polycyclic aromatic hydrocarbons (PAHs) using human blood and plasma analysis", Journal of Chromatography, B, 878:1753–1760.

Pleil JD, Stiegel MA, Madden MC, and Sobus JR, 2011a. **"**Heat Map Visualization of Complex Environmental and Biomarker Measurements", Chemosphere, on-line: doi:10.1016/j.chemosphere.2011.03.017

Pleil JD, Stiegel MA, Sobus JR, Liu Q, and Madden MC, 2011b. **"**Interpreting the breath biomarker exposome for environmental and intelligence applications", Journal of Breath Research 5:03714 (9pp).

Rappaport SM and Smith MT, 2011. "Environment and disease risk", Science 330(6003):460-461.

Rappaport SM, 2011. "Implications of the exposome for exposure science", Journal of Exposure Science and Environmental Epidemiology 21:5–9.

Rioux JD and Abbas AK, 2005. "Paths to understanding the genetic basis of autoimmune disease", Nature 435:584-589.

Sawyer K, Samet JD, Ghio AJ, Pleil JD, and Madden MC, 2008. "Responses measured in the exhaled breath of human volunteers acutely exposed to ozone and diesel exhaust", Journal of Breath Research, 2:037019, 9pp.

Smith MT and Rappaport SM. 2009. "Building Exposure Biology Centers to Put the E intro "GxE" Interaction Studies", Environmental Health Perspectives 117(8): A334-A335.

Sobus JR, Pleil JD, McClean MD, Herrick RF, and Rappaport SM, 2010. "Biomarker variance component estimation for exposure surrogate selection and toxicokinetic inference", Toxicology Letters 199(3):247-253.

Sobus JR, Tan YM, Pleil JD, Sheldon LS, 2011. "A biomonitoring framework to support exposure and risk assessments", Science of the Total Environment, (in press)

Wild CP, 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14(8):1847-50.

Wild CP, 2011. How much of a contribution do exposures experienced between conception and adolescence make to the burden of cancer in adults? Cancer Epidemiol Biomarkers Prev 20(4):580-581.