

Title: Biomarker variance component estimation for exposure surrogate selection and toxicokinetic inference

Authors: Jon R. Sobus^{a*}, Joachim D. Pleil^a, Michael D. McClean^b, Robert F. Herrick^c, and Stephen M. Rappaport^d

Affiliations:

^aHuman Exposure and Atmospheric Sciences Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA, sobus.jon@epa.gov, pleil.joachim@epa.gov;

^bDepartment of Environmental Health, Boston University School of Public Health, 715 Albany Street, Talbot Building, Boston, MA 02118, USA, mmclean@bu.edu;

^cExposure, Epidemiology and Risk Program, Department of Environmental Health, Harvard School of Public Health, Landmark Center, 401 Park Drive, Boston, MA 02215, USA, herrick@hohp.harvard.edu;

^dDivision of Environmental Health Sciences, School of Public Health, University of California, 50 University Hall #7360, Berkeley, CA 94720-7360, USA, srappaport@berkeley.edu;

Corresponding Author: *Correspondence to: Jon R. Sobus, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA, Mail Code: E205-04, sobus.jon@epa.gov, phone: (919) 943-5900, fax: (919) 541-0905

Abstract

Biomarkers are useful exposure surrogates given their ability to integrate exposures through all routes and to reflect interindividual differences in toxicokinetic processes. Also, biomarker concentrations tend to vary less than corresponding environmental measurements, making them less-biasing surrogates for exposure. In this article, urinary PAH biomarkers (namely, urinary naphthalene [U-Nap]; urinary phenanthrene [U-Phe]; 1-hydroxypyrene [1-OH-Pyr]; and 1-, (2+3)-, 4-, and 9-hydroxyphenanthrene [1-, (2+3)-, 4-, and 9-OH-Phe]) were evaluated as surrogates for exposure to hot asphalt emissions using data from 20 road-paving workers. Linear mixed-effects models were used to estimate the within- and between-person components of variance for each urinary biomarker. The ratio of within- to between-person variance was then used to estimate the biasing effects of each biomarker on a theoretical exposure-response relationship. Mixed models were also used to estimate the amounts of variation in Phe metabolism to individual OH-Phe isomers that could be attributed to Phe exposure (as represented by U-Phe concentrations) and covariates representing time, hydration level, smoking status, age, and body mass index. Results showed that 1-OH-Phe, (2+3)-OH-Phe, and 1-OH-Pyr were the least-biasing surrogates for exposure to hot asphalt emissions, and that effects of hydration level and sample collection time substantially inflated bias estimates for the urinary biomarkers. Mixed-model results for the individual OH-Phe isomers showed that between 63% and 82% of the observed biomarker variance was collectively explained by Phe exposure, the time and day of sample collection, and the hydration level, smoking status, body mass index, and age of each worker. By difference, the model results also showed that, depending on the OH-Phe isomer, a maximum of 6% to 23% of the total biomarker variance was attributable to differences in unobserved toxicokinetic processes between the workers. Therefore, toxicokinetic processes are probably less influential on urinary biomarker variance than are exposures and observable covariate effects. The methods described in this analysis should be considered for the selection and interpretation of biomarkers as exposure surrogates in future exposure investigations.

Keywords: biomarkers, exposure, variance components, PAHs

1.0 Introduction

Biomonitoring has emerged as an important tool for research in the field of exposure science. Indeed, biological measurements offer several theoretical advantages over traditional environmental measurements of toxicants in air, water, on the skin, and in food (Lin et al., 2005). First, biomarkers integrate exposures from all sources and routes thereby reducing the need to monitor all environmental sources separately. Second, biomarker concentrations tend to vary less than the corresponding environmental levels measured from day to day. This reduction in intra-subject variability of biomarker levels, relative to environmental levels, reduces the number of measurements required to precisely assess exposures. And third, biomarkers reflect interindividual variations in toxicokinetic processes (i.e., uptake, metabolism, and elimination) that may be important to the health impact of exposures.

We previously explored the utility of urinary biomarkers for characterizing multi-route exposures to polycyclic aromatic hydrocarbons (PAHs), a large class of compounds which includes numerous carcinogens, in a longitudinal investigation of 20 road-paving workers exposed to emissions from hot-mix asphalt (Sobus et al., 2009a; Sobus et al., 2009b). Specifically, we examined urinary levels of three surrogate PAHs and their hydroxylated metabolites, namely those of naphthalene (Nap; a 2-ring PAH), phenanthrene (Phe; a 3-ring PAH), and pyrene (Pyr; a 4-ring PAH), as functions of total PAH concentrations in airborne particulate matter and on dermal patches. Since as many as nine urine samples were collected per subject, these data were also used to estimate within-subject and between-subject components of variance for each PAH biomarker in this population. Results from linear mixed-effects models showed significant effects on urinary biomarker levels of both airborne and dermal patch measurements of total PAHs. As such, urinary naphthalene (U-Nap), urinary phenanthrene (U-Phe), 1-hydroxypyrene (1-OH-Pyr), and the sum of 5 monohydroxylated metabolites of phenanthrene (OH-Phe) were identified as useful biomarkers of combined air and dermal exposures to PAHs (Sobus et al., 2009a).

Environmental and biological measurements of exogenous substances vary greatly over time within a given subject. This within-subject variability leads to uncertainty in ‘true’ exposure levels (long-term means) which leads, in turn, to biased predictions of exposure-response relationships in epidemiologic studies (Armstrong and Oakes, 1982; Lin et al., 2005). Since a goal of environmental epidemiology is to accurately link environmental exposures with

observed health outcomes, it is desirable to select measures of exposure that are less variable over time and, therefore, less biasing surrogates for true exposure levels. Using data from our study of road-paving workers, we compare repeated observations of U-Nap, U-Phe, 1-OH-Pyr, and 1-, 2-, 3-, 4-, and 9-OH-Phe to select the least-biasing surrogate(s) for exposures to hot asphalt emissions.

Because biomarkers reflect interpersonal differences in toxicokinetic processes, they can also be useful for characterizing variability in rates of metabolism of toxic substances across populations. To explore this idea, we use data from our road-pavers study to investigate relationships between unmetabolized Phe (U-Phe) and the monohydroxylated metabolites of Phe (OH-Phe isomers) in the same urine samples. Specifically, we use linear mixed-effects models to regress the level of a given isomer of OH-Phe on the corresponding level of U-Phe while adjusting for covariates. Since U-Phe takes into account the absorption and distribution of Phe, the unexplained between-subject variability in the levels of the OH-Phe isomers can be used to infer the variation caused by metabolism and elimination. Also, regression coefficients for covariates in the models can be used to quantify the effects of potential modifiers (e.g., smoking status, BMI, and age) on human metabolism of Phe.

Although we examine urinary PAH biomarkers among a group of road-paving workers, the methods are sufficiently general to be applied to other volatile organic compounds that are excreted in urine as both parent compounds and metabolites. As such, these methods can inform the selection and interpretation of biomarkers for many toxic substances.

2.0 Methods

2.1 Study design and measurements of urinary PAH analytes

Subjects included 20 male road-paving workers residing in the Greater Boston area of the United States who were recruited with informed consent under protocols approved by committees for human-subjects research at participating institutions. Workers' exposures to PAHs were evaluated over three consecutive work days, starting at the beginning of the workweek following a work-free weekend, using breathing-zone air and dermal patch measurements of PAHs as described in detail elsewhere (McClellan et al., 2004a; McClellan et al., 2004b). Urine samples were collected immediately after each workshift, at bedtime, and in the

morning following each workday, using sterilized polypropylene containers. Up to nine repeated urine samples were collected from each worker, starting with the postshift void on the first workday. Urine samples were stored at -20°C for approximately seven years prior to analysis for urinary PAH analytes; a previous examination of these data showed that appreciable effects of prolonged storage time on the urinary analyte levels were unlikely (Sobus et al., 2009b). Concentrations of U-Nap and U-Phe were determined using headspace-solid phase microextraction coupled with gas chromatography-mass spectrometry (Sobus et al., 2009c), and concentrations of 1-OH-Pyr, 1-OH-Phe, (2+3)-OH-Phe, 4-OH-Phe, and 9-OH-Phe were determined using solid-phase extraction coupled with liquid chromatography-tandem mass spectrometry (Onyemauwa et al., 2009). The estimated limit of detection was 0.40 ng/L for U-Nap and U-Phe (Sobus et al., 2009c), and the estimated limits of quantitation were 2.0 ng/L for (2+3)-OH-Phe, and 5.0 ng/L for 1-OH-Pyr, 1-OH-Phe, 4-OH-Phe, and 9-OH-Phe (Onyemauwa et al., 2009). The estimated coefficients of variation for all urinary PAH analytes were within a range of 0.053 to 0.27 (Onyemauwa et al., 2009; Sobus et al., 2009c). Creatinine measurements were determined using a colorimetric assay, and each subject's height, weight, age, and smoking status were obtained by questionnaire (Sobus et al., 2009a). Summary statistics for each of these covariates are given in (Sobus et al., 2009a).

2.2 Selecting the least-biasing biomarker of exposure

Exposures are measured with error using either environmental or biomarker concentrations as surrogates for the true exposure levels. The simplest way to quantify this biasing effect of exposure measurement error is to consider an individual-based study where a given exposure surrogate is measured repeatedly for each person in a sample, and a continuous health outcome (e.g. respiratory function or DNA adducts) is also known for each person (Rappaport and Kupper, 2008). Assuming a simple straight-line relationship between the logged health outcome and logged exposure surrogate in the population, the true slope of the exposure-response relationship is defined as β_{true} . However, because the exposure surrogate is measured with error, the straight-line slope of the *estimated* exposure-response relationship is not β_{true} but rather is β_{est} . The ratio of β_{est} to β_{true} determines the amount of attenuation bias for a given exposure surrogate under this simple model, as given by the following relationship (Rappaport and Kupper, 2008):

$$b = \frac{\beta_{\text{est}}}{\beta_{\text{true}}} = \left(1 + \frac{\lambda}{n_i}\right)^{-1}, \quad (1)$$

where the amount of bias = $(1 - b)$ (e.g. if $b = 0.8$ then bias = 0.2, indicating 20% attenuation bias); the variance ratio $\lambda = \frac{\sigma_w^2}{\sigma_b^2}$, for σ_w^2 and σ_b^2 representing the respective within-person and between-person variance components (of logged exposure values); and n_i represents the number of repeated measurements for each person. From (1) it is seen that, with measurement error, the estimated exposure-response relationship is attenuated because $\beta_{\text{est}} < \beta_{\text{true}}$ and that that b varies between 0 and 1. Furthermore, for a given value of n_i , the amount of bias increases with $\lambda = \frac{\sigma_w^2}{\sigma_b^2}$ indicating that the variance ratio dictates the amount of attenuation bias in a given study. Since each exposure surrogate has its own value of λ , variance ratios can be compared to determine the least-biasing measure of exposure, i.e. the surrogate with the smallest λ .

In our study of different urinary biomarkers of PAH exposure, values of the variance ratios are estimated as $\hat{\lambda} = \frac{\hat{\sigma}_w^2}{\hat{\sigma}_b^2}$, where $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ are the estimates of σ_w^2 and σ_b^2 obtained by linear mixed models of the logged biomarker measurements (as discussed in the next section). Also, we define \hat{b} as the estimate of b derived from (1) after substituting $\hat{\lambda}$ for λ . Finally, we use (1) to estimate the number of repeated measurements needed to limit attenuation bias at a given value (Rappaport and Kupper, 2008) for a particular biomarker of exposure, that is

$$n_i = \left\lceil \frac{\hat{b}}{(1 - \hat{b})} \hat{\lambda} \right\rceil. \quad (2)$$

For example, if $\hat{b} = 0.8$ and $\hat{\sigma}_w^2 = \hat{\sigma}_b^2$, then $\hat{\lambda} = 1$ and $n_i = 4$. This indicates that 4 repeated measurements of a biomarker of exposure with an estimated variance ratio of 1 would be required to limit attenuation bias to 20%.

2.3 Mixed effects models

Restricted maximum likelihood (REML) estimates of σ_w^2 and σ_b^2 were determined for each urinary PAH analyte using linear mixed effects models (Proc MIXED of SAS version 9.1).

Null models (containing only random effects and a global mean) were first created for each analyte, followed by reduced models that included fixed effects for urinary creatinine concentration (an inversely proportional surrogate of hydration level), time of sample collection, and day of sample collection, as shown in equation (3).

$$Y_{ij} = \ln(X_{ij}) = \beta_0 + \beta_1 \text{CREATININE}_{ij} + \beta_2 \text{TIME}_{ij} + \beta_3 \text{DAY}_{ij} + b_i + \varepsilon_{ij} \quad (3)$$

for $j = 1, 2, \dots, n_i$ measurements of the i th individual; and
for $i = 1, 2, \dots, 20$ individuals.

Here, X_{ij} represents the concentration of a urinary PAH analyte (ng/L) for the j th measurement of the i th person and Y_{ij} is the natural logarithm of the individual measurement X_{ij} . The coefficient β_0 is the model intercept, and β_1 , β_2 , and β_3 represent the coefficients for the fixed effects of CREATININE (creatinine concentration [$\ln(\text{g/l})$]), TIME (categorical variable for postshift, bedtime, or morning), and DAY (categorical variable for day 1, day 2, or day 3), respectively. In equation (3), b_i is the random effect for the i th person and ε_{ij} is the random error for the j th measurement of the i th person. It is assumed that b_i and ε_{ij} are independent random variables and that $b_i \sim N(0, \sigma_b^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_w^2)$. Fixed effects in the final reduced models were selected using backwards stepwise elimination at a significance level of $p \leq 0.1$. A compound symmetry covariance structure was used in each of the null and reduced models as this structure generally yielded the lowest Akaike's information criterion (AIC) and Bayesian information criterion (BIC) values compared to other tested structures (i.e., autoregressive [AR(1)] and heterogeneous autoregressive [ARH(1)]).

Additional linear mixed-effects models were developed for the individual isomers of OH-Phe to investigate the effects of exposure to Phe (as indicated by levels of U-Phe) on analyte levels after adjusting for covariate effects. Equation (4) shows the full mixed model for the individual isomers of OH-Phe.

$$Y_{ij} = \ln(X_{ij}) = \beta_0 + \beta_1 \text{CREATININE}_{ij} + \beta_2 \text{TIME}_{ij} + \beta_3 \text{DAY}_{ij} + \beta_4 \text{UPHE}_{ij} + (\beta_5 \text{UPHE}_{ij} \times \text{TIME}_{ij}) + (\beta_6 \text{UPHE}_{ij} \times \text{DAY}_{ij}) + \beta_7 \text{SMOKER}_i + \beta_8 \text{BMI}_i + \beta_9 \text{AGE}_i + b_i + \varepsilon_{ij} \quad (4)$$

for $j = 1, 2, \dots, n_i$ measurements of the i th individual; and
for $i = 1, 2, \dots, 20$ individuals.

Here, as in equation (3), b_i and ε_{ij} are independent random variables, β_0 is the model intercept, and β_1 , β_2 and β_3 represent the coefficients for the fixed effects of CREATININE, TIME and DAY, respectively. The coefficients β_4 , β_5 , β_6 , β_7 , β_8 , and β_9 correspond to the fixed effects of $UPHE_{ij}$ (the [logged] concentration of U-Phe for j^{th} measurement of the i^{th} person), $UPHE_{ij} \times TIME_{ij}$ (the interaction between UPHE and TIME), $UPHE_{ij} \times DAY_{ij}$ (the interaction between UPHE and DAY), $SMOKER_i$ (the smoking status of the i^{th} worker, where nonsmoker = 0), BMI_i (the body mass index [kg/m^2] of the i^{th} worker), and AGE_i (the age in years of the i^{th} worker), respectively. A compound symmetry covariance matrix was used in each full model as this structure generally yielded the lowest values of AIC and BIC. All covariates were maintained in the final full models for ease of interpretation, but were determined to be significant at $p \leq 0.10$.

Estimates of within- and between-subject fold-ranges (i.e., ${}_w\hat{R}_{0.95}$ and ${}_b\hat{R}_{0.95}$, respectively) were determined for each OH-Phe isomer under equation (4) according to Rappaport (Rappaport, 1991) where: ${}_w\hat{R}_{0.95} = e^{3.92\hat{\sigma}_w}$ and ${}_b\hat{R}_{0.95} = e^{3.92\hat{\sigma}_b}$. The value ${}_w\hat{R}_{0.95}$ represents the estimated fold-range containing 95% of the biomarker measurements for a typical person in the population investigated, and the value ${}_b\hat{R}_{0.95}$ represents the fold-range containing 95% of the mean biomarker levels across all persons. Intraclass correlation coefficients (ICC)

were also estimated for each OH-Phe isomer under equation (4), where $ICC_{full} = \frac{\hat{\sigma}_{b_{full}}^2}{\hat{\sigma}_{w_{full}}^2 + \hat{\sigma}_{b_{full}}^2}$;

estimates of ICC_{full} represent the ratio of between-subject biomarker variance to total observed variance under each full model. Finally, the percents of total variance (where total estimated variance $\hat{\sigma}_Y^2 = \hat{\sigma}_w^2 + \hat{\sigma}_b^2$) from the null models explained by fixed effects in the reduced and full

models were determined as follows: $\% \text{explained}_{red} = \left[\frac{(\hat{\sigma}_{Y_{null}}^2 - \hat{\sigma}_{Y_{red}}^2)}{\hat{\sigma}_{Y_{null}}^2} \right] \times 100$ and

$$\% \text{explained}_{full} = \left[\frac{(\hat{\sigma}_{Y_{null}}^2 - \hat{\sigma}_{Y_{full}}^2)}{\hat{\sigma}_{Y_{null}}^2} \right] \times 100, \text{ respectively (Burstyn et al., 2000; Egeghy et al., 2005;$$

Sobus et al., 2009a).

3.0 Results

3.1 Estimates of variance components and attenuation bias for urinary PAH biomarkers

Table 1 shows geometric mean (GM) urinary analyte levels measured in samples of postshift, bedtime, and morning urine, along with variance component and %Bias estimates from the null models and from the reduced models after adjusting for CREATININE, TIME, and DAY. Overall, the GM levels of the monohydroxylated analytes were approximately one to two orders of magnitude greater than those of the unmetabolized analytes (i.e., U-Nap and U-Phe). Furthermore, the GM levels of each analyte followed a rank order of postshift > bedtime > morning, indicating a rapid uptake and elimination of Nap, Phe, and Pyr during each workday.

Estimates of the variance ratio λ from the null models (i.e., $\hat{\lambda}_{null}$) were generally greater than one, indicating that the estimated within-person variance components were larger than the estimated between-person variance components, particularly for U-Nap ($\hat{\lambda}_{null}=3$), 4-OH-Phe ($\hat{\lambda}_{null}=4$), and U-Phe ($\hat{\lambda}_{null}=8$) (see Table 1). Based on estimated values of λ_{null} , an attenuation bias of 18% or less would be expected when using the levels of 9-OH-Phe, (2+3)-OH-Phe, 1-OH-Phe, or 1-OH-Pyr as surrogates for ‘true’ exposure levels; attenuation biases of 24%, 33%, and 47% would be expected using levels of U-Nap, 4-OH-Phe, and U-Phe, respectively.

Significant effects of CREATININE ($p < 0.0001$) and TIME ($p < 0.002$) were observed in the reduced models for each urinary analyte, and a significant effect of DAY ($p < 0.1$) was observed for all analytes except U-Phe and 9-OH-Phe. Positive regression coefficients for CREATININE (ranging from 0.765 to 1.38) indicate higher concentrations of urinary analytes with decreased hydration levels, and the regression coefficients for TIME support the observed rank order of postshift levels > bedtime levels > morning levels in all cases. In instances of a significant DAY effect, analyte levels were always lowest on the first sampling day, suggesting slight accumulation of individual biomarkers at the beginning of the workweek.

Figure 1 shows residual values over time of U-Phe (Figure 1A) and (2+3)-OH-Phe (Figure 1B) from the null models, and Figure 2 shows residual values of the same analytes after adjusting for significant fixed effects. In Figures 1A and 1B, the group mean levels deviate from zero over time (e.g., mean postshift residual levels are greater than zero and morning levels are below zero); this variability highlights significant time trends for which adjustments were not made in the null models. Figures 2A and 2B show random variation of the residual values about zero at all time points, reflecting the appropriate incorporation of fixed time effects into the reduced models. While significant time effects were observed for each analyte, residuals of U-Phe and (2+3)-OH-Phe are shown in Figures 1 and 2 as examples of the extent to which measurement error (or within-worker variance) differs between selected biomarkers. Indeed, the residuals of repeated observations for individual workers in Figures 1A and 2A (U-Phe) are considerably more varied than those in Figures 2A and 2B (2+3-OH-Phe).

After adjusting for significant fixed effects, the estimate of λ from the reduced model (i.e., $\hat{\lambda}_{red}$) was observed to be lower than that of λ_{null} for each analyte (see Table 1), highlighting a greater than proportional decrease in $\hat{\sigma}_{w_{null}}^2$ to $\hat{\sigma}_{w_{red}}^2$ than in $\hat{\sigma}_{b_{null}}^2$ to $\hat{\sigma}_{b_{red}}^2$. A median decrease of 60% (with a range of 51% to 69%) was observed from $\hat{\sigma}_{w_{null}}^2$ to $\hat{\sigma}_{w_{red}}^2$ indicating that CREATININE, TIME and DAY contributed, in large part, to the variation in spot urine measurements within individual workers. In contrast, only about a $\pm 10\%$ difference was observed between $\hat{\sigma}_{b_{null}}^2$ and $\hat{\sigma}_{b_{red}}^2$ for U-Nap, U-Phe, 1-OH-Phe, (2+3)-OH-Phe, and 4-OH-Phe, indicating that the average levels of these analytes between individual workers were not appreciably affected by CREATININE, TIME or DAY.

The lowest levels of $\hat{\lambda}_{red}$ were observed for 1-OH-Phe, (2+3)-OH-Phe, and 1-OH-Pyr, with estimated values ranging from approximately 0.5 – 0.7 (Table 1). Assuming nine repeated observations for each worker, the %Bias_{red} for these respective analytes was determined to be 5%, 6%, and 7%. The calculated values of %Bias_{red} for the remaining analytes (i.e., U-Nap, 4-OH-Phe, 9-OH-Phe, and U-Phe) were $\geq 13\%$.

3.2 Mixed model results for individual isomers of OH-Phe

Results of the full mixed models for the individual isomers of OH-Phe are shown in Table 2. Highly significant ($p < 0.0001$) positive effects of UPHE and CREATININE were observed in each model indicating elevated levels of OH-Phe isomers with increased levels of UPHE and CREATININE. Significant interaction effects of both UPHE×TIME and UPHE×DAY were observed for 1-OH-Phe, (2+3)-OH-Phe, and 9-OH-Phe ($p < 0.10$), indicating a change in the (logged) linear relationship between UPHE and these OH-Phe isomers over time. These significant interaction effects, and the main effects of DAY and TIME for individual analytes, are probably artifacts of the different elimination rates of U-Phe and the OH-Phe isomers (Sobus et al., 2009b). After adjusting for the main effects of CREATININE, UPHE, TIME, and DAY, and the interaction effects of UPHE×TIME and UPHE×DAY, a significant smoking effect was observed for 1-OH-Phe and 9-OH-Phe. Interestingly, smokers produced *higher* levels of 9-OH-Phe compared to non-smokers, whereas smokers produced *lower* levels of 1-OH-Phe compared to nonsmokers. Only 9-OH-Phe and (2+3)-OH-Phe were significantly affected by BMI ($p \leq 0.10$), with lower analyte levels observed in high BMI workers. The covariate ‘AGE’ was not significantly associated with the levels of OH-Phe isomers in any of the full models.

Figure 3 shows the percents of total estimated biomarker variance in the null models explained by fixed effects in the reduced and full models (obtained using estimates of $\hat{\sigma}_{w_{null}}^2$ and $\hat{\sigma}_{b_{null}}^2$, $\hat{\sigma}_{w_{red}}^2$ and $\hat{\sigma}_{b_{red}}^2$, and $\hat{\sigma}_{w_{full}}^2$ and $\hat{\sigma}_{b_{full}}^2$, from Tables 1 and 2). The fixed effects of CREATININE, TIME and DAY from the reduced models collectively explained 39-58% of the biomarker variance in the null models, whereas all fixed effects from the full models collectively explained 63-82% of the biomarker variance in the null models. These results indicate that the majority of the observed OH-Phe isomer variance was explained by the road workers’ estimated exposure levels and by observed covariate effects.

By difference (i.e., subtracting the explained variance from 100%), the estimates of unexplained variance in the full models for 1-OH-Phe, (2+3)-OH-Phe, 4-OH-Phe, and 9-OH-Phe were 37%, 32%, 37%, and 18%, respectively. The ICC estimates from the full models (shown in Table 2) suggest that about 60% of the unexplained variance in 1-OH-Phe and (2+3)-OH-Phe levels, and about 30% of the unexplained variance in 4-OH-Phe and 9-OH-Phe levels, was observed between subjects. Considering these estimates of unexplained variance and of ICCs,

the between-subject components of variance were 23%, 21%, 12%, and 6% of the total unexplained variance in the respective levels of 1-OH-Phe, (2+3)-OH-Phe, 4-OH-Phe, and 9-OH-Phe. These results suggest that, depending on the OH-Phe isomer, a maximum of 6% to 23% of the total biomarker variance was attributable to differences in unobserved toxicokinetic processes between the workers.

4.0 Discussion

This analysis explored the theoretical advantages of biomonitoring using existing data from an observational study of 20 road-paving workers. Null, reduced and full linear mixed effects models were constructed to highlight the attenuation bias associated with individual biomarkers of PAHs emitted from hot asphalt, and to quantify the amount of biomarker variance attributable to Phe exposure, covariate influence, and unknown toxicokinetic processes.

The biasing potential of selected air and biomarker measurements as exposure surrogates have been evaluated in other studies (Egeghy et al., 2005; Fustinoni et al., 2010; Liljelind et al., 2003; Lin et al., 2005; Rappaport et al., 1995). Notably, Lin and colleagues estimated variance components of individual analytes using approximately 12,000 repeated air and biomarker measurements compiled from over 100 different datasets (Lin et al., 2005). Results from their analysis suggested that a given biomarker measurement is likely a less-biasing surrogate for exposure than is a typical air measurement. Additionally, the results of Lin et al. suggested an inverse relationship between the biasing-potential of a given biomarker and its residence time in the body (Lin et al., 2005). We have previously estimated the biological half-lives of U-Phe and OH-Phe to be approximately 8 and 14 h, respectively (Sobus et al., 2009b). Therefore, our results support this earlier observed inverse relationship, as estimates of λ_{null} and λ_{red} were both larger for U-Phe than for the individual OH-Phe isomers (see Table 1). Furthermore, our results show that the effects of physiological damping (i.e., diminished measurement error with increased residence time (Rappaport and Kupper, 2008)) on biomarker variance are observable even when comparing short-term biomarkers (residence time on the order of hours or days), and not just short- and intermediate-term, or short- and long-term biomarkers.

To help interpret the results of the mixed models used in this analysis, we present three hypothetical exposure assessment scenarios for the 20 road-paving workers; we assume that the results of these three hypothetical assessments would be analogous to those of our null, reduced

and full mixed models. Under the null models we present hypothetical scenario #1, in which approximately nine biomarker measurements were made of each of the 20 workers, with each measurement made at a randomly selected time (i.e., postshift, bedtime, or morning), on a randomly selected workday (i.e., day 1, day 2, or day 3), in a randomly selected workweek. Under the reduced models we present hypothetical scenario #2, in which approximately nine biomarker measurements were made of each of the 20 workers (one measurement per worker per week), with all measurements made at approximately the same time of day (either postshift, bedtime, or morning), on the same day of the workweek (either day 1, day 2, or day 3), over nine randomly selected workweeks. Additionally, all 20 workers under scenario #2 were equally hydrated at the time of sample collection. Under the full models we present hypothetical scenario #3, in which approximately nine biomarker measurements were made of each of these 20 workers (one measurement per worker per week), with measurements made at approximately the same time of day, on the same day of the workweek, over nine randomly selected workweeks. Additionally, all 20 workers under scenario #3 were equally hydrated at the time of sample collection, had approximately the same exposure profile over time, had the same BMI, were non-smokers, and were the same age.

For each of the three hypothetical scenarios, $\hat{\sigma}_b^2$ represents the differences in average biomarker levels across workers, and $\hat{\sigma}_w^2$ represents the differences in individual biomarker measurements for any given worker over time. Under scenario #1 (null models), the magnitude of $\hat{\sigma}_b^2$ likely reflects differences between workers in average exposure levels, personal characteristics (e.g., smoking status, BMI and age), and toxicokinetic processes, whereas the magnitude of $\hat{\sigma}_w^2$ likely reflects changes in exposure levels for a given worker, the time and day of sample collection, and the hydration levels of the workers at the time of sample collection. Results from the null models (shown in Table 1) suggested that the ratio of $\hat{\sigma}_{w_{null}}^2$ to $\hat{\sigma}_{b_{null}}^2$ (i.e., $\hat{\lambda}_{null}$) was lowest for 1-OH-Pyr, 1-OH-Phe, and (2+3)-OH-Phe, indicating that these metabolites were likely the least-biasing surrogates for exposure to hot asphalt emissions. Attenuation bias estimates for these three analytes were between 12-15% whereas those for the other analytes ranged from 18-47%. If it were necessary to restrict attenuation bias to 10% in a prospective exposure study (under scenario #1), about 11-14 measurements per worker would be needed of either 1-OH-Pyr, 1-OH-Phe, or (2+3)-OH-Phe vs. about 70 measurements per worker of U-Phe,

the biomarker with the largest estimated level of λ_{null} . This disparity in sample size highlights the importance of estimating variance components in exposure studies.

Under scenario #2 (reduced models), the magnitude of $\hat{\sigma}_b^2$ again likely reflects differences between workers in average exposure levels, personal characteristics (e.g., smoking status, BMI, age), and toxicokinetic processes, and the magnitude of $\hat{\sigma}_w^2$ reflects changes in exposure levels for a given worker. However $\hat{\sigma}_w^2$ is *not* influenced by the time or day of sample collection nor the hydration levels of the workers. Estimates of σ_w^2 from the reduced models were considerably lower than those observed in the null models, indicating that the majority of the biomarker variance for any given worker could be explained by TIME, DAY, and CREATININE. As expected, estimates of σ_b^2 from these models were very similar (in all but two cases) to those observed in the null models, indicating that the variability in the *average* biomarker levels across all workers was not appreciably affected by TIME, DAY, and CREATININE. Taken together, the results of $\hat{\lambda}_{null}$ and $\hat{\lambda}_{red}$ (shown in table 1) suggest a considerable reduction in the biomarker attenuation bias under scenario #2. For example, only 5-6 measurements of 1-OH-Phe, (2+3)-OH-Phe, or 1-OH-Pyr would be required to achieve a 10% attenuation bias under scenario #2 vs. 11-14 measurements of these biomarkers under scenario #1. The results of this comparison indicate that, given increased observations of influential covariates, fewer measurements are needed to obtain a reliable estimate of exposure.

The earlier work of Lin et al. evaluated fixed time effects (seasonal effects, weekday, effects, and linear trends) on biomarker levels using mixed models (Lin et al., 2005). Results of that analysis indicated a larger impact of time on $\hat{\sigma}_w^2$ compared to $\hat{\sigma}_b^2$, and suggested that omission of an important time effect increases values of $\hat{\lambda}$. Since our results for the urinary PAH biomarkers are consistent with the earlier work of Lin et al., we hereby confirm the need to adjust for time effects when comparing exposure surrogates to a continuous health outcome. Furthermore, we establish a clear need to adjust for changes in the hydration level of individuals when considering urinary biomarker measurements as exposure surrogates.

Under scenario #3 (full models), it is assumed that little or no biomarker variance stems from differences in exposure levels, the time and day of sample collection, hydration levels, smoking status, BMI, or age. Thus, the magnitude of $\hat{\sigma}_b^2$ likely reflects variations in

unidentified toxicokinetic processes between workers and the magnitude of $\hat{\sigma}_w^2$ likely reflects minor sources of measurement error (including laboratory assay error) and unresolved covariate effects. Obviously the assumptions under scenario #3 are impossible to replicate in observational exposure studies. However, human chamber experiments provide a unique platform with which to evaluate biomarker measurements in a controlled environment. As such, a retrospective analysis was recently performed of the variability in blood and breath biomarker measurements of trichloroethylene (TCE), methyl tertiary butyl ether (MTBE), and tertiary butyl alcohol (TBA; a phase I metabolites of MTBE), using data from two human chamber experiments (Pleil, 2009). In the original chamber studies, nominally healthy and non-smoking subjects were exposed to equal concentrations of either TCE or MTBE through the inhalation, ingestion, or dermal routes (Pleil et al., 1998; Prah et al., 2004), and blood and breath biomarkers of TCE, MTBE, and TBA were measured at specific time points. Because the exposures were well controlled in each of these experiments, the magnitude of biomarker variance between subjects at each time point was thought to be a function of internal biological processes. Following from this assumption, the effects of biological processes on between-subject biomarker variance was quantified at each time point using fold-range estimates (i.e., the ratio of the upper and lower bounds of the 95% confidence intervals, which was determined using the GM and GSD of the log-normal biomarker distributions (Pleil, 2009)). Results showed that the between-subject fold-range estimates were very similar across all analytes and biological media, and ranged from 2.3 to 6.2 (averaged across all time points) (Pleil, 2009). That is, 95% of the observed biomarker levels for the study subjects were within an approximate 2- to 7-fold-range, given identical exposure conditions.

In the full mixed models used for this investigation, UPHE was selected as an exposure surrogate to control for variations in Phe exposures between- and within-workers. Therefore, the residual between-subject variance estimates from the full models here are directly comparable to the variance estimates from the human chamber studies. Results from our full models (Table 2) showed that the between-subject fold-range estimates for the individual OH-Phe isomers were between 3.09 and 6.71. That is, 95% of the mean biomarker levels for the 20 workers were within an approximate 3 to 7-fold-range, after controlling for differences in Phe exposures, hydration level, and observed covariates. This result is essentially the same as observed from the chamber study analysis of Pleil (Pleil, 2009) for different volatile organic compounds (TCE,

MTBE, and TBA) and suggests that intersubject variability in unobserved toxicokinetic processes may be relatively consistent across volatile compounds. It is important to note, however, that the within-subject fold-range estimates (i.e., ${}_w\hat{R}_{0.95}$) from our full models (Table 2) were very similar to the respective estimates of ${}_bR_{0.95}$, indicating that 95% of the individual biomarker measurements for any given worker were within an approximate 4 to 8 fold-range (after controlling for fixed effects). These results likely point to some laboratory measurement error, some unaddressed covariate influence, and differences in toxicokinetic parameters between U-Phe and the OH-Phe isomers for which adjustments were not made in the full models.

The final objective of this analysis was to estimate the amount of observed OH-Phe biomarker variance attributable to Phe exposures, covariate influence, and unidentified toxicokinetic processes. Fixed effects in the reduced models collectively explained about 40-60% of the observed biomarker variance, whereas those from the full models explained about 60-80% of the observed biomarker variance. Therefore, for each OH-Phe isomer, a majority of the observed variance was attributable to changes in exposure levels and covariate conditions. However, given the substantial differences in unexplained variance *between isomers* in the reduced and full models, additional analyses of the effects of toxicokinetic processes on human PAH metabolism are warranted. Furthermore, the differential effects of SMOKER and BMI on individual OH-Phe isomers (see Table 2) point to some aspects of PAH metabolism that are affected by personal characteristics; these effects on metabolism should be examined further in future observational studies of PAH-exposed subjects.

5.0 Conclusions

We have identified 1-OH-Phe, (2+3)-OH-Phe, and 1-OH-Pyr as the least-biasing surrogates for exposure to hot-asphalt emissions, and have shown that changes in hydration level and the timing of sample collection can substantially inflate bias estimates for urinary PAH biomarkers. Furthermore, we have shown that toxicokinetic processes are probably less influential on urinary biomarker variance than are exposure and covariate effects. Finally, we used individual isomers of OH-Phe to demonstrate that exposure and covariate effects can vary substantially depending on the isomeric form of the biomarker. Therefore, we conclude that the estimation of biomarker variance components in exposure studies is necessary for selecting the least-biasing exposure surrogate(s), for quantifying the effects of toxicokinetic processes on

biomarker variance, and for resolving the differential effects of covariates on exposure-biomarker or biomarker-response relationships. The methods used for this analysis should be considered for the selection and interpretation of biomarkers as exposure surrogates in future investigations.

6.0 Acknowledgements

The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here. It has been subjected to Agency review and approved for publication. Additional funding was provided by the National Cancer Institute (grant R01 CA74413-03) and the National Institutes of Health; National Institutes of Environmental Health Sciences (training grant T32ES07018, research grant P42ES05948, and center grant P30ES10126). The authors are grateful for the expert advice and assistance from Peter Egeghy, Rogelio Tornero-Velez, Ronald Williams, Donald Whitaker, Marsha Morgan, Roy Fortmann, and Linda Sheldon of U.S. EPA, Suramya Waidyanatha of NIEHS, David Kim of Syngenta Crop Protection, Inc., and Lawrence Kupper of the University of North Carolina at Chapel Hill.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

Table 1. Estimated geometric means (ng/L), variance components, and attenuation bias for urinary biomarkers of naphthalene, phenanthrene, and pyrene (where $n_i \leq 9$).

	1-OH-Pyr	1-OH-Phe	(2+3)-OH-Phe	9-OH-Phe^a	U-Nap	4-OH-Phe	U-Phe^a
Postshift GM (GSD)	2050 (2.63)	1190 (2.32)	3530 (2.25)	2070 (2.47)	89.9 (2.14)	483 (2.24)	69.1 (2.62)
Bedtime GM (GSD)	1080 (3.93)	822 (2.92)	2300 (2.74)	948 (3.58)	48.0 (2.28)	203 (2.58)	31.5 (3.44)
Morning GM (GSD)	689 (3.71)	434 (2.52)	1080 (2.13)	514 (3.16)	36.7 (2.20)	141 (2.59)	13.2 (3.52)
<i>n</i>	154	154	154	154	161	154	161
$\hat{\sigma}_{w_{null}}^2$ (SE)	0.903 (0.110)	0.598 (0.073)	0.594 (0.072)	1.05 (0.129)	0.574 (0.068)	0.883 (0.107)	1.59 (0.189)
$\hat{\sigma}_{b_{null}}^2$ (SE)	0.767 (0.286)	0.463 (0.175)	0.388 (0.150)	0.527 (0.216)	0.198 (0.089)	0.198 (0.101)	0.203 (0.132)
$\hat{\lambda}_{null}$	1.18	1.29	1.53	1.99	2.90	4.46	7.83
%Bias_{null}	12	13	15	18	24	33	47
$\hat{\sigma}_{w_{red}}^2$ (SE)	0.381 (0.047)	0.220 (0.027)	0.187 (0.023)	0.402 (0.050)	0.279 (0.034)	0.358 (0.045)	0.772 (0.093)
$\hat{\sigma}_{b_{red}}^2$ (SE)	0.583 (0.205)	0.430 (0.149)	0.355 (0.124)	0.256 (0.101)	0.211 (0.082)	0.219 (0.087)	0.225 (0.106)
$\hat{\lambda}_{red}$	0.654	0.512	0.527	1.57	1.32	1.64	3.43
%Bias_{red}	7	5	6	15	13	15	28

^areduced model included a fixed effect for CREATININE and TIME, but not DAY.
GM, geometric mean; GSD, geometric standard deviation; SE, standard error.

Table 2. Results from full linear mixed-effects models (equation (4)) for individual isomers of OH-Phe ($n = 154$).

Parameters	1-OH-Phe		(2+3)-OH-Phe		4-OH-Phe		9-OH-Phe	
	Estimate (SE)	<i>p</i> -value	Estimate (SE)	<i>p</i> -value	Estimate (SE)	<i>p</i> -value	Estimate (SE)	<i>p</i> -value
Fixed effects								
Intercept [ln(ng/L)]	5.83 (0.800)	<0.0001	7.11 (0.734)	<0.0001	3.40 (0.694)	0.0002	6.42 (0.573)	<0.0001
CREATININE [ln(g/L)]	0.637 (0.101)	<0.0001	0.522 (0.086)	<0.0001	0.597 (0.127)	<0.0001	0.662 (0.109)	<0.0001
TIME		0.4		0.1		0.3		0.05
postshift	-0.169 (0.322)		-0.346 (0.274)		-0.049 (0.412)		-0.021 (0.356)	
bedtime	-0.340 (0.240)		-0.424 (0.204)		-0.490 (0.308)		-0.627 (0.267)	
morning	0 (ref)		0 (ref)		0 (ref)		0 (ref)	
DAY		0.006		0.001		0.2		0.04
day 3	-0.131 (0.240)		0.384 (0.204)		-0.371 (0.311)		-0.517 (0.270)	
day 2	0.565 (0.223)		0.720 (0.190)		0.218 (0.289)		0.155 (0.251)	
day 1	0 (ref)		0 (ref)		0 (ref)		0 (ref)	
U-PHE [ln(ng/L)]	0.200 (0.062)	<0.0001	0.191 (0.052)	<0.0001	0.261 (0.079)	<0.0001	0.408 (0.069)	<0.0001
U-PHE×TIME		0.02		<0.0001		0.4		0.06
U-Phe×postshift	0.139 (0.083)		0.258 (0.070)		0.114 (0.106)		0.073 (0.092)	
U-Phe×bedtime	0.202 (0.073)		0.272 (0.062)		0.129 (0.094)		0.192 (0.081)	
U-Phe×morning	0 (ref)		0 (ref)		0 (ref)		0 (ref)	
U-PHE×DAY		0.01		0.01		0.1		0.06
U-Phe×day 3	0.073 (0.066)		-0.052 (0.056)		0.145 (0.085)		0.139 (0.074)	
U-Phe×day 2	-0.110 (0.061)		-0.152 (0.052)		-0.005 (0.078)		-0.017 (0.068)	
U-Phe×day 1	0 (ref)		0 (ref)		0 (ref)		0 (ref)	
SMOKER		0.009		0.2		0.2		0.06
yes	-0.719 (0.270)		-0.361 (0.250)		-0.273 (0.218)		0.338 (0.177)	
no	0 (ref)		0 (ref)		0 (ref)		0 (ref)	
BMI (kg/m ²)	-0.024 (0.026)	0.3	-0.036 (0.024)	0.1	0.016 (0.021)	0.5	-0.047 (0.017)	0.007
AGE (years)	0.009 (0.010)	0.4	0.005 (0.009)	0.6	0.008 (0.008)	0.3	-2E ⁻⁴ (0.006)	1.0
Random effects								
$\sigma_{w_{full}}^2$	0.156 (0.020)	<0.0001	0.112 (0.014)	<0.0001	0.264 (0.034)	<0.0001	0.200 (0.025)	<0.0001
$\sigma_{b_{full}}^2$	0.236 (0.092)	0.005	0.205 (0.080)	0.005	0.130 (0.061)	0.02	0.083 (0.041)	0.02
$wR_{0.95}$	4.70		3.71		7.49		5.77	
$bR_{0.95}$	6.71		5.90		4.11		3.09	
ICC _{full}	0.602		0.647		0.330		0.293	

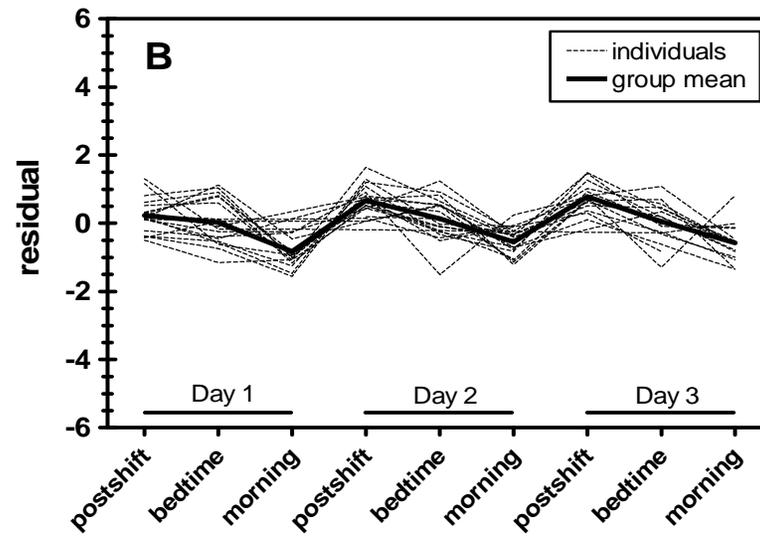
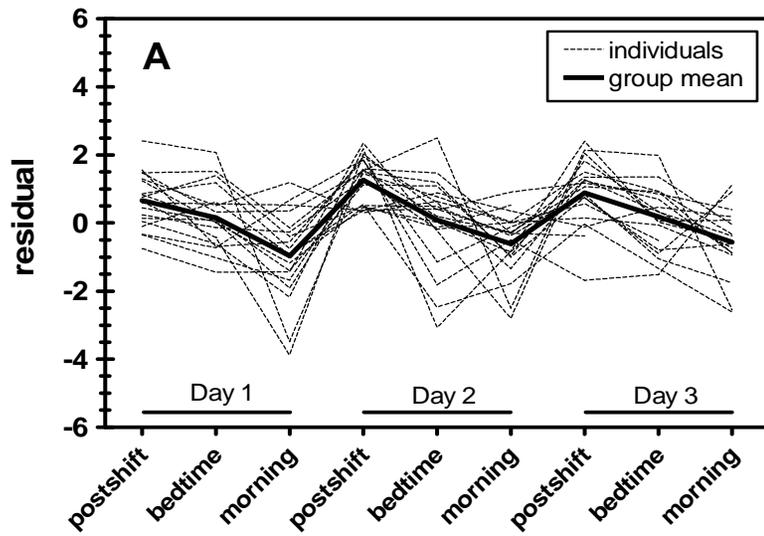


Figure 1. Residuals for U-Phe (A) and (2+3)-OH-Phe (B) over time under the null models.

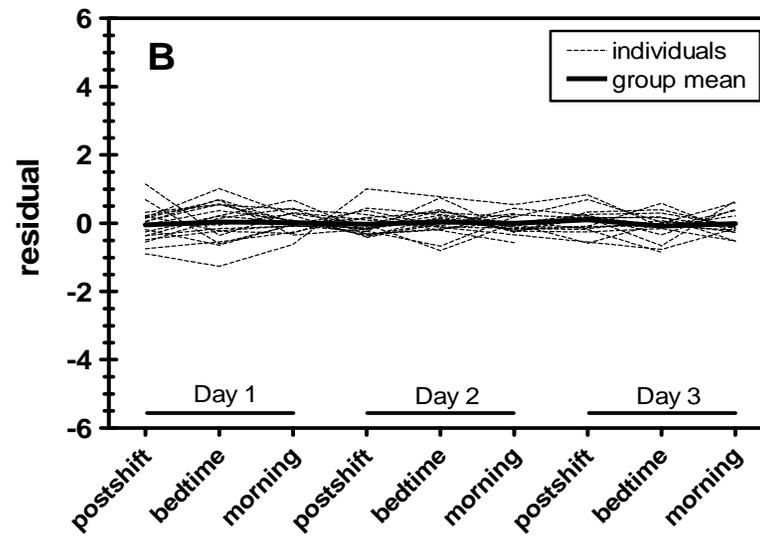
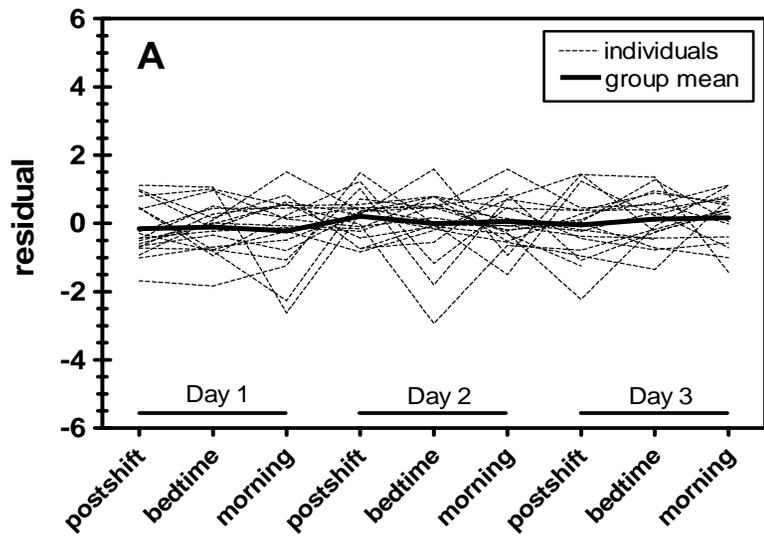


Figure 2. Residuals for U-Phe (A) and (2+3)-OH-Phe (B) over time under the reduced models [equation (3)].

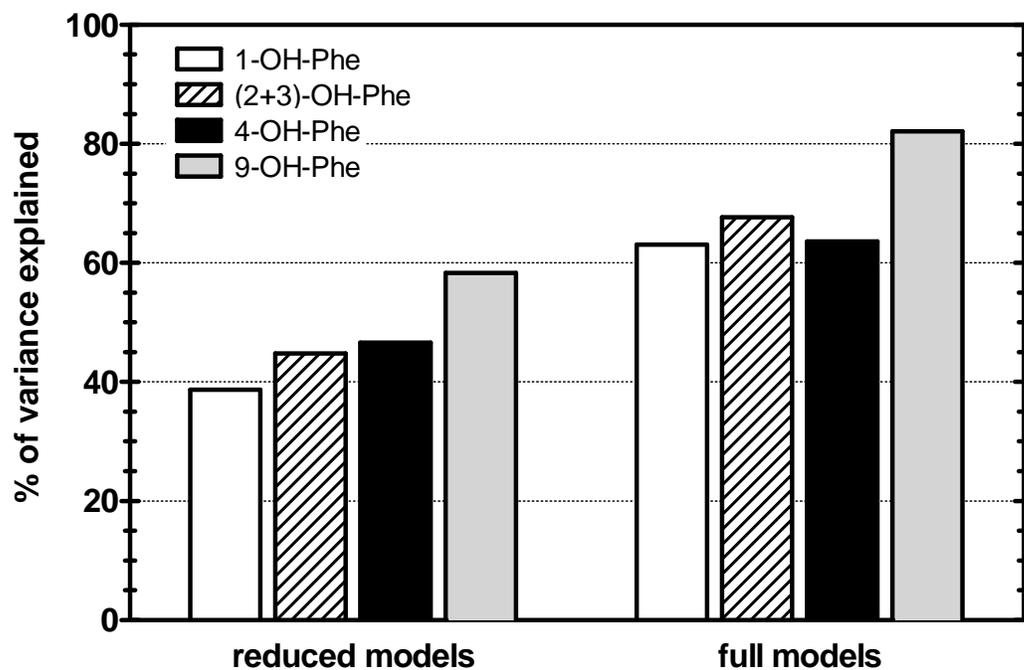


Figure 3. Percent of total OH-Phe isomer variance from null models explained by fixed effects in reduced [equation (3)] and full [equation (4)] mixed models.

References

- Armstrong, B.G., Oakes, D., 1982. Effects of approximation in exposure assessments on estimates of exposure- response relationships. *Scand J Work Environ Health* 8 Suppl 1, 20-23.
- Burstyn, I., Kromhout, H., Kauppinen, T., Heikkila, P., Boffetta, P., 2000. Statistical modelling of the determinants of historical exposure to bitumen and polycyclic aromatic hydrocarbons among paving workers. *Ann Occup Hyg* 44, 43-56.
- Egeghy, P.P., Quackenboss, J.J., Catlin, S., Ryan, P.B., 2005. Determinants of temporal variability in NHEXAS-Maryland environmental concentrations, exposures, and biomarkers. *J Expo Anal Environ Epidemiol* 15, 388-397.
- Fustinoni, S., Manini, P., Campo, L., De Palma, G., Andreoli, R., Mutti, A., Bertazzi, P.A., Rappaport, S.M., 2010. Assessing variability and comparing short-term biomarkers of styrene exposure using a repeated measurements approach. *Toxicol Lett* 192, 40-44.
- Liljelind, I., Rappaport, S., Eriksson, K., Andersson, J., Bergdahl, I.A., Sunesson, A.L., Jarvholm, B., 2003. Exposure assessment of monoterpenes and styrene: a comparison of air sampling and biomonitoring. *Occup Environ Med* 60, 599-603.
- Lin, Y.S., Kupper, L.L., Rappaport, S.M., 2005. Air samples versus biomarkers for epidemiology. *Occup Environ Med* 62, 750-760.
- McClean, M.D., Rinehart, R.D., Ngo, L., Eisen, E.A., Kelsey, K.T., Herrick, R.F., 2004a. Inhalation and dermal exposure among asphalt paving workers. *Ann Occup Hyg* 48, 663-671.
- McClean, M.D., Rinehart, R.D., Ngo, L., Eisen, E.A., Kelsey, K.T., Wiencke, J.K., Herrick, R.F., 2004b. Urinary 1-hydroxypyrene and polycyclic aromatic hydrocarbon exposure among asphalt paving workers. *Ann Occup Hyg* 48, 565-578.
- Onyemauwa, F., Rappaport, S.M., Sobus, J.R., Gajdosova, D., Wu, R., Waidyanatha, S., 2009. Using liquid chromatography-tandem mass spectrometry to quantify monohydroxylated metabolites of polycyclic aromatic hydrocarbons in urine. *J Chromatogr B Analyt Technol Biomed Life Sci* 877, 1117-1125.
- Pleil, J.D., 2009. Influence of systems biology response and environmental exposure level on between-subject variability in breath and blood biomarkers. *Biomarkers* 14, 560-571.
- Pleil, J.D., Fisher, J.W., Lindstrom, A.B., 1998. Trichloroethene levels in human blood and exhaled breath from controlled inhalation exposure. *Environ Health Perspect* 106, 573-580.
- Prah, J., Ashley, D., Blount, B., Case, M., Leavens, T., Pleil, J., Cardinali, F., 2004. Dermal, oral, and inhalation pharmacokinetics of methyl tertiary butyl ether (MTBE) in human volunteers. *Toxicol Sci* 77, 195-205.
- Rappaport, S.M., 1991. Assessment of long-term exposures to toxic substances in air. *Ann Occup Hyg* 35, 61-121.
- Rappaport, S.M., Kupper, L.L., 2008. Quantitative exposure assessment. Stephen M. Rappaport, El Cerrito, CA.
- Rappaport, S.M., Symanski, E., Yager, J.W., Kupper, L.L., 1995. The relationship between environmental monitoring and biological markers in exposure assessment. *Environ Health Perspect* 103 Suppl 3, 49-53.
- Sobus, J.R., McClean, M.D., Herrick, R.F., Waidyanatha, S., Nylander-French, L.A., Kupper,

- L.L., Rappaport, S.M., 2009a. Comparing urinary biomarkers of airborne and dermal exposure to polycyclic aromatic compounds in asphalt-exposed workers. *Ann Occup Hyg* 53, 561-571.
- Sobus, J.R., McClean, M.D., Herrick, R.F., Waidyanatha, S., Onyemauwa, F., Kupper, L.L., Rappaport, S.M., 2009b. Investigation of PAH biomarkers in the urine of workers exposed to hot asphalt. *Ann Occup Hyg* 53, 551-560.
- Sobus, J.R., Waidyanatha, S., McClean, M.D., Herrick, R.F., Smith, T.J., Garshick, E., Laden, F., Hart, J.E., Zheng, Y., Rappaport, S.M., 2009c. Urinary naphthalene and phenanthrene as biomarkers of occupational exposure to polycyclic aromatic hydrocarbons. *Occup Environ Med* 66, 99-104.