AMERICAN JOURNAL OF INDUSTRIAL MEDICINE 50:112–121 (2007)

Evaluating Measurement Error in Estimates of Worker Exposure Assessed in Parallel by Personal and Biological Monitoring

Elaine Symanski, PhD,[1]* Nicole M.H. Greeson, MS, CIH,[2] and Wenyaw Chan, PhD[3]

Background *While studies indicate that the attenuating effects of imperfectly measured exposure can be substantial, they have not had the requisite data to compare methods of assessing exposure for the same individuals monitored over common time periods.*
Methods *We examined measurement error in multiple exposure measures collected in parallel on 32 groups of workers. Random-effects models were applied under both compound symmetric and exponential correlation structures. Estimates of the within- and between-worker variances were used to contrast the attenuation bias in an exposure-response relationship that would be expected using an individual-based exposure assessment for different exposure measures on the basis of the intra-class correlation coefficient (ICC).*
Results *ICC estimates ranged widely, indicative of a great deal of measurement error in some exposure measures while others contained very little. There was generally less attenuation in the biomarker data as compared to measurements obtained by personal sampling and, among biomarkers, for those with longer half-lives. The interval ICC estimates were oftentimes wide, suggesting a fair amount of imprecision in the point estimates. Ignoring serial correlation tended to over estimate the ICC values.*
Conclusions *Although personal sampling results were typically characterized by more intra-individual variability than inter-individual variability when compared to biological measurements, both types of data provided examples of exposure measures fraught with error. Our results also indicated substantial imprecision in the estimates of exposure measurement error, suggesting that greater emphasis needs to be given to studies that collect sufficient data to better characterize the attenuating effects of an error-prone exposure measure.* Am. J. Ind. Med. 50:112–121, 2007. © *2006 Wiley-Liss, Inc.*

KEY WORDS: occupational exposure; epidemiology; measurement error; intra-class correlation coefficient

INTRODUCTION

Intra-individual variability in exposure can induce error in exposure assessment and thereby adversely affect epidemiologic studies by diminishing measures of effect and the power to detect significant effects [Thomas et al., 1993; Armstrong, 1998; Loomis and Kromhout, 2004]. If exposures are estimated with little error, effects on study results are minimal. On the other hand, effects may be substantial for measures of exposure subject to considerable error, as would be the case for exposures that vary greatly over time relative to inter-individual differences in true

[1]Division of Environmental and Occupational Health Sciences, University of Texas School of Public Health at Houston, Houston, Texas
[2]Occupational and Environmental Safety Office, Duke University, Durham, North Carolina
[3]Division of Biostatistics, University of Texas School of Public Health at Houston, Houston, Texas

Contract grant sponsor: The National Institute for Occupational Safety and Health (NIOSH), U.S. Centers for Disease Control and Prevention; Contract grant number: R03 OH007834.

*Correspondence to: Elaine Symanski, Environmental and Occupational Health Sciences, University of Texas School of Public Health at Houston, 1200 Herman Pressler Drive, RAS 642, Houston, Texas 77030. E-mail: Elaine.Symanski@uth.tmc.edu

Accepted 14 November 2006
DOI 10.1002/ajim.20422. Published online in Wiley InterScience (www.interscience.wiley.com)

exposure levels. Thus, in studies that rely on an individual-based approach to estimate workers' average exposure levels (rather than on a group-based approach in which average exposure is estimated for an occupational group of workers and assigned to all members of the group), it is desirable to select an exposure measure that maximizes differences among workers' exposures relative to the variation that occurs across sampling periods to minimize effects due to measurement error.

Studies to assess measurement error in monitoring data have been carried out on workers exposed to styrene in the reinforced plastics industry [Rappaport et al., 1995; Symanski et al., 2001; Liljelind et al., 2003], mercury in the chloralkali industry [Symanski et al., 2000], and dust in the carbon black manufacturing industry [van Tongeren et al., 1997], the construction industry [Tjoe et al., 2004], and sawmills [Teschke et al., 2004]. Taken together, these studies indicate that attenuation effects of imperfectly measured exposure can be quite substantial. Few studies, however, have had the requisite data to make comparisons between different methods of assessing exposure [Rappaport et al., 1995; Symanski et al., 2000; Symanski et al., 2001; Liljelind et al., 2003; Lin et al., 2005], and none employed a study design to insure that the same workers were monitored over common periods using multiple sampling methods. Thus, we carried out a study to quantify measurement error in multiple measures of exposure collected in parallel on groups of workers exposed to different contaminants from a broad cross-section of industries. Given such error in exposure measures, a secondary objective was to contrast the degree of attenuation in an exposure-response relationship that would be expected using an individual-based exposure assessment approach for different exposure measures.

MATERIALS AND METHODS

Estimating Exposure Effects in the Presence of Exposure Measurement Error

Based upon the classical measurement error model [Fuller, 1987] and under the assumption that the underlying exposure distribution is lognormal, the model equation for the exposure-response relationship becomes:

$$Y_i = \alpha + \beta \ln X_i + \nu_i \quad \text{for } i = 1, 2, \ldots, k \text{ workers} \qquad (1)$$

where Y_i is the continuous response for the i-th individual, X_i is the true mean exposure for the i-th individual, α is the intercept, β is the slope, and ν_i is the response error term. In model (1), we assume that $\ln X_i's$ are independent and distributed normally with mean μ_X and variance σ_B^2 and that ν_i is distributed normally with a mean of 0 and variance σ_ν^2. Given that $X_i{=}x_i$, Y_i is distributed normally with mean $\alpha{+}\beta$ $\ln x_i$ and variance σ_Y^2.

When X_i is unobservable, we rely on Z_i, which represents an imperfect measure of X_i where:

$$\ln Z_i = \ln X_i + \varepsilon_i \qquad (2)$$

In Equation 2 it is assumed that the measurement error, ε_i, is distributed normally with a mean of 0 and variance σ_W^2. It directly follows that $\ln Z_i$ is distributed normally with a mean of μ_X and a variance of σ_Z^2 where $\sigma_Z^2 = \sigma_B^2 + \sigma_W^2$. It is noteworthy to mention that the relationship between Z_i and X_i includes a multiplicative error term, that is, $Z_i = X_i e^{\varepsilon_i}$.

It has been shown [Thomas et al., 1993] that the ratio of the ''observed'' coefficient (β') (obtained using the observed Z_i values) to the true coefficient (β) can be expressed as follows:

$$\frac{\beta'}{\beta} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \qquad (3)$$

Thus, regressing Y_i on the imperfect measure $\ln Z_i$, instead of the true value $\ln X_i$, yields a slope coefficient β' that is attenuated. The factor $\sigma_B^2/(\sigma_B^2 + \sigma_W^2)$ represents the degree of attenuation of the true slope coefficient β.

Similarly, for each X_i, we may observe n randomly collected measurements of the surrogate measure on j occasions (i.e., $n Z_{ij}'s$) and compare the true exposure-response regression with the regression of Y_i's on $\overline{\ln Z_i}$ (where $\overline{\ln Z_i} = \sum_{j=1}^{n} \ln Z_{ij}/n$). Since $E(\ln X_i | \ln Z_{i1}, \ldots, \ln Z_{in}) = E(\ln X_i | \overline{\ln Z_i})$, the ratio of the observed coefficient to the true coefficient becomes:

$$\frac{\beta'}{\beta} = \frac{\sigma_B^2}{\sigma_B^2 + \frac{\sigma_W^2}{n}} \qquad (4)$$

As before, regressing Y_i on the logarithm of measured exposure $\overline{\ln Z_i}$, instead of the log-scale of the true exposure X_i, yields a slope coefficient β' that is attenuated. Under expectation, the factor $\sigma_B^2/(\sigma_B^2 + \sigma_W^2/n)$ represents the degree of attenuation of the slope coefficient β.

Estimating Measurement Error in Multiple Measures of Exposures Collected in Tandem on Workers From Different Industries

The database

The database contributing measurements analyzed in this study has been described elsewhere [Symanski and Greeson, 2002]. Briefly, biological monitoring data were abstracted from 53 studies reported in the peer-reviewed literature that examined workers' exposures in myriad industries worldwide. To be eligible for inclusion, repeated measurements must have been collected on workers

to allow for the quantification of the between- and within-person sources of variation in exposure. Groups of workers were classified by location (plant) with one restriction (a group of drycleaners from multiple sites). Personal air sampling results, if available, were compiled as well. In addition to groups of workers on whom both personal and biological monitoring had been conducted, multiple biomarkers of exposure were available on some groups. In total, the database contains 121 sets of data, comprised of 6,174 biological and personal exposure measurements collected on 577 workers from 55 groups. In addition to the exposure measurements, details about the industry, the airborne contaminant, the sampling protocol, and the worker were abstracted from the original studies. For each biomarker of exposure, information about the half-life of the contaminant or its metabolite was also recorded.

For inclusion in the current study, data were eliminated on the basis of the following criteria: (1) a single biological monitoring method was used to evaluate exposure, (2) biological measurements were collected prior to or at the start of the work-shift (if comparisons were to be made to personal monitoring data), (3) fewer than five workers had been monitored, (4) reported use of respirators was inconsistent among workers with both personal and biological monitoring data, or (5) workers came from more than one workplace (drycleaners). To enhance comparability among multiple measures of exposures collected on groups of workers, data were further examined to insure that the measurements were collected concurrently on the same workers on common days of sampling. In instances where data were missing on some workers (i.e., measurements were available on one exposure measure but not for another) or where the periods of sampling for different exposure measures did not exactly coincide, monitoring data on specific workers or individual measurements were omitted to obtain sets of data that were balanced across exposure measures on the basis of both the numbers of workers sampled and days monitored. Following these exclusions, the data were re-examined to exclude groups with fewer than five workers or less than 10 measurements.

The model to quantify measurement error

To assess the attenuation bias for each measure of exposure, a one-way random-effects model was applied following natural logarithmic transformation of the personal or biological measurements of exposure. Under the assumption that the variance and covariance functions do not depend on the time of measurement, we considered two covariance structures for the log-transformed measurements collected on the same individual: (1) a compound symmetric structure and (2) an exponential correlation structure. Under compound symmetry (CS), it is assumed that the covariance between measurements collected on the same worker is the same irrespective of the time interval separating them [$\mathrm{Cov}(\ln Z_{ij}, \ln Z_{ij'}) = \sigma_B^2 \quad \text{for } j \neq j'$]. It directly follows that the correlation between measurements on the same person can be expressed as [$\mathrm{Corr}(\ln Z_{ij}, \ln Z_{ij'}) = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$ for $j \neq j'$], which is also known as the intra-class correlation coefficient (ICC). We expected in some cases that the correlation between repeated measurements would decrease as a function of the time interval between them, as might occur for a biomarker with a long half-life in which a series of measurements collected on the same worker were collected over a relatively short period. Thus, an exponential (EXP) correlation structure (in which the correlation is dependent upon the time interval between measurements) was also applied. Here, the covariance function is $\mathrm{Cov}(\ln Z_{ij}, \ln Z_{ij'}) = \sigma_B^2 \rho^{\tau_{jj'}} \sigma_W^2 \quad \text{for } j \neq j'$ and where $\rho^{\tau_{jj'}}$ is the correlation between measurements separated by the interval $\tau_{jj'}[\tau_{jj'} = |t_j - t_{j'}|]$. Under this structure, the correlation decays toward zero as the time separation between measurements increases (which represents a generalization of the first-order autotoregressive [AR(1)] structure for equally spaced data). Under both correlation structures, the within-worker (σ_W^2) and between-worker (σ_B^2) variances were estimated using the method of restricted maximum likelihood (REML) implemented with the PROC MIXED procedure from the SAS System Software (Version 9.1, Cary, NC).

Since in a simple linear regression it is the relative magnitude of the variances that provides an indication of the bias in the slope coefficient due to measurement error in the exposure measure (refer to Equation 3 above), estimated values of the ICC under CS were computed as [$\hat{\sigma}_B^2/(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)$]. Exact 95% confidence limits on the ICC were also obtained according to the methods outlined by Searle et al. [1992], which rely on functions of the mean squares associated with the sources of variation in the one-way random effects model. Under assumptions of normality, the REML and ANOVA estimators for the variance components are identical in the balanced case *except* when the ANOVA estimators yield a negative estimate (since REML estimates are always non-negative). In this instance, the ANOVA estimators of the variances were obtained to allow for the interval estimation of the ICCs. Negative point or interval estimates were set to 0. Estimated variances using the ANOVA method of estimation were obtained with PROC NESTED in SAS.

Finally, in expanding upon the work of Lin et al. [2005], we derived an expression for the ICCs for serially correlated, irregularly spaced measurements as follows:

$$\frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \frac{\hat{\sigma}_W^2}{n}\left(1 + \frac{2}{n}\sum_{j=2}^{n}\sum_{j'=1}^{j-1}\hat{\rho}^{\tau_{jj'}}\right)} \tag{5}$$

Equation 5 was applied to all data sets with evidence of serial correlation ($P < 0.05$) under the model with an exponential correlation structure.

RESULTS

After applying the criteria for inclusion in the study, data were available on 32 groups of workers with multiple measures of exposure representing a total of 77 datasets—each group represents workers from a common workplace and each data set compiles monitoring data for a single exposure measure. Fourteen groups of workers were sampled via both personal and biological monitoring and 18 groups were assessed with multiple biological measures of exposure. Table I provides a listing of groups of workers on whom both personal and biological monitoring had been conducted with information on the industry, the exposure measure, the duration of the monitoring period, the time of sampling, and numbers of workers and measurements. Groups 3 and 7 had personal samples on two related exposures (alkyl lead and inorganic lead and PAHs and benzo(a)pyrene, respectively), and five groups (groups 1, 2, 3, 10, and 13) had

TABLE I. Groups of Workers Whose Exposures Were Assessed by Air and Biological Monitoring (n = 14 groups)[a]

Group	N	k	Duration of sampling	Industry	Type	Time	Units
1	27	9	1 week	Hospital sterilizer unit	Environmental ethylene oxide	Postshift	mg/m^3
					Blood ethylene oxide	Postshift	μg/L
					Alveolar ethylene oxide	Postshift	mg/m^3
2	25	5	1 week	Dry alkaline battery manufacturing	Inorganic mercury	8:00 AM–2:00 PM	μg/m^3
					Blood mercury	2:00 PM	μg/100 ml
					Urinary mercury	2:00 PM	μg/g creatinine
3	80	5	1 month	Alkyl lead manufacturing	Alkyl lead	During shift	μg/m^3
					Inorganic lead	During shift	μg/m^3
					Urinary lead	During shift	μg/L
					Urinary ALA	During shift	mg/100 ml
4	20	5	1 week	Plastic boat manufacturing	Styrene	During shift	Ppm
					Alveolar styrene	Postshift	ppm
5	15	5	1 week	Carbon black manufacturing	Dust	During shift	mg/m^3
					Urinary 1-hydroxypyrene	Postshift	μmol/mol creatinine
6	40	20	1 year	Lead-acid battery manufacturing	Lead	During shift	μg/m^3
					Blood lead	During shift	μg/100 ml
7	20	10	1 week	Coke manufacturing	Benzo[a]pyrene	During shift	μg/m^3
					Total PAHs	During shift	μg/m^3
					Urinary 1-hydroxypyrene	Postshift	ng/ml
8	16	8	1 week	Coke manufacturing	Benzo[a]pyrene	During shift	μg/m^3
					Urinary 1-hydroxypyrene	Postshift	ng/ml
9	25	5	1 week	Pulp and paper industry machine manufacturing	Water-soluble chromium	During shift	mg/m^3
					Urinary chromium	4:00 PM	μg/g creatinine
10	16	8	10 months	Plastic boat manufacturing	Styrene	During shift	ppm
					Urinary mandelic acid	Postshift	mg/g creatinine
					Urinary phenylglyoxylic acid	Postshift	mg/g creatinine
11	10	5	1 week	Viscose rayon plant	Carbon disulfide	During shift	ppm
					Urinary TTCA	Postshift	mg/g creatinine
12	18	6	1 week	Creosote impregnation plant	Pyrene	During shift	μg/m^3
					Urinary 1-hydroxypyrene	Endshift	μmol/mol creatinine
13	12	6	7 months	Plastics manufacturing	Styrene	During shift	mg/m^3
					Blood styrene	Midshift	mg/L
					Urinary mandelic acid	Postshift	mmol/mol creatinine
14	63	7	1 month	Chloralkali processing plant	Inorganic mercury	During shift	μg/m^3
					Blood mercury	Postshift	nmol/L

[a]Abbreviations used: N, no. of measurements; k, no. of workers; ALA, δ-aminolevulinic acid; PAHs, polycyclic aromatic hydrocarbons; TTCA, 2-thiothiazolidine-4-carboxylic acid.

TABLE II. Groups of Workers Whose Exposures Were Assessed by Multiple Biomarkers (n = 18 groups)[a]

Group	N	k	Duration of sampling	Industry	Type	Time	Units
15	22	11	1 week	Copper smelting	Urinary inorganic arsenic	Postshift	μg/L
					Urinary methylarsonic acid	Postshift	μg/L
					Urinary dimethylarsinic acid	Postshift	μg/L
					Urinary trimethylarsenic compounds	Postshift	μg/L
16	30	10	1 week	Rotogravure printing	Alveolar air toluene	Postshift	μg/L
					Blood toluene	Postshift	mg/kg
					Urinary hippuric acid	Postshift	g/L
17	26	13	2 years	Cadmium pigment manufacturing	Urinary cadmium		μg/L
					Blood cadmium		μg/100 ml
18	124	44	5 months	Sawmilling	Urinary tetrachlorophenol		ppb
					Urinary pentachlorophenol		ppb
19	12	6	1 week	Plastic boat manufacturing	Urinary mandelic acid + phenylglyoxylic acid	4:00 PM	mmol/h
					Urinary mandelic acid + phenylglyoxylic acid	9:00 PM	mmol/h
20	66	12	5 years	Chromium plating	Urinary chromium	During shift	μg/g creatinine
					Sister Chromatid exchanges/cell (mean)	During shift	
21	12	6	1 week	Polyurethane manufacturing	Plasma 2,4-TDA	3:30 PM	ng/ml
					Plasma 2,6-TDA	3:30 PM	ng/ml
22	40	5	2 months	Forestry	Blood arsenic	Postshift	ppm
					Urinary arsenic	24 hr	μg/24 hr
23	30	5	1 month	Alkyl lead manufacturing	Urinary lead	During shift	mg/L
					Urinary ALA	During shift	mg/100 ml
24	25	5	1 month	Alkyl lead manufacturing	Blood lead	During shift	μg/100 g
					Urinary lead	During shift	μg/L
					Urinary ALA	During shift	mg/100 ml
25	22	11	2 years	Plastic boat manufacturing	Blood styrene	Postshift	mg/L
					Urinary mandelic acid	Postshift	mg/g creatinine
					Urinary phenylglyoxylic acid	Postshift	mg/g creatinine
26	10	5	1 week	Smelting	Urinary inorganic As + MMA + DMA	Preshift	μg/g creatinine
					Urinary inorganic As + MMA + DMA	Preshift	μg/L
27	14	7	2 months	Plastics manufacturing	Blood styrene	Midshift	mg/L
					Urinary mandelic acid	Postshift	mmol/mol creatinine
					06-styrene-guanine lymphocyte adducts	Midshift	adducts/10–8 dNp
28	32	8	1 week	Nickel refining	Urinary nickel	Postshift	μmol/L
					Urinary nickel	Postshift	mmol/mol creatinine
29	18	9	6 months	Chemical manufacturing	Blood mercury		μg/L
					Urinary mercury		μg/g creatinine
30	106	22	2 months	Tobacco farming	Plasma cholinesterase	PM	μmols/ml/min
					Red blood cell cholinesterase	PM	μmols/ml/min
31	55	11	2 months	Tobacco farming	Plasma cholinesterase	PM	μmols/ml/min
					Red blood cell cholinesterase	PM	μmols/ml/min
32	62	14	2 months	Tobacco farming	Plasma cholinesterase	PM	μmols/ml/min
					Red blood cell cholinesterase	PM	μmols/ml/min

[a]Abbreviations used: N, no. of measurements; k, no. of workers; ALA, δ-aminolevulinic acid; TDA, 2,4-toluenediamine; MMA, methylarsonic acid; DMA, dimethylarsonic acid.

been monitored using two biological measures of exposure. Table II provides a breakdown of groups on whom only biological monitoring had been conducted. Among these workers, groups 15, 16, 24, 25, and 27 had been evaluated with three or more biomarkers. Measurements on groups 24, 25, and 27 were collected in the same workplace as groups 3, 10, and 13, respectively (Table I), but not on the same set of workers or over the same sampling intervals. Characteristics

TABLE III. Characteristics of the Air-Biological Exposure Measures and Multiple Biomarkers Databases

	Air-biomarker database	Multiple biomarker database
Number of groups	14	18
Number of data sets	35	42
Industrial sector[a]		
Agriculture, hunting and forestry	—	1 group
Mining and quarrying	—	3 groups
Manufacturing	13 groups	14 groups
Health and social work	1 group	—
Type of contaminant (exposure)		
Gas/vapor	5 groups	6 groups
Aerosol	6 groups	12 groups
Combination	3 groups	—
Number of workers		
Total	104	204
Range (per group)	5–20	5–44
Average (across groups)	7	11
Median (across groups)	6	10
Number of measurements		
Total	1,034	1,547
Range (per data set)	10–80	10–124
Average (across data sets)	30	37
Median (across data sets)	20	26
Sampling interval		
1 week or less	9 groups	6 groups
1 month–3 months	2 groups	7 groups
5 months–1 year	3 groups	2 groups
>1 year	—	3 groups
Residence time of biomarker		
Short (<7 days)	12 data sets	21 data sets
Intermediate (1–4 weeks)	5 data sets	11 data sets
Long (>4 weeks)	2 data sets	9 data sets

[a]Based on the International Standard Industrial Classification (ISIC) coding [United Nations, 1990].

TABLE IV. Point and Interval Estimates of the Intra-Class Correlation Coefficients (ICCs) for Workers' Exposures Assessed in Parallel by Air and Biological Monitoring (Based Upon the Model With Compound Symmetry)

Group	Type	ICC[a]	95% CI
1	Environmental ethylene oxide	0.000	—[b]
	Blood ethylene oxide	0.000	(0.000, 0.273)
	Alveolar ethylene oxide	0.000	—[b]
2	Inorganic mercury	0.757	(0.426, 0.966)
	Blood mercury	0.798	(0.494, 0.972)
	Urinary mercury	0.985	(0.947, 0.998)
3	Alkyl lead	0.204	(0.043, 0.722)
	Inorganic lead	0.000	(0.000, 0.294)
	Urinary lead	0.634	(0.353, 0.937)
	Urinary ALA	0.649	(0.368, 0.941)
4	Styrene	0.000	(0.000, 0.619)
	Alveolar styrene	0.000	(0.000, 0.628)
5	Dust	0.228	(0.000, 0.839)
	Urinary 1-hydroxypyrene	0.738	(0.271, 0.965)
6	Lead	0.993	(0.984, 0.997)
	Blood lead	0.943	(0.864, 0.977)
7	Benzo[a]pyrene	0.000	(0.000, 0.167)
	Total PAHs	0.000	(0.000, 0.143)
	Urinary 1-hydroxypyrene	0.399	(0.000, 0.805)
8	Benzo[a]pyrene	0.495	(0.000, 0.871)
	Urinary 1-hydroxypyrene	0.636	(0.000, 0.913)
9	Water-soluble chromium	0.658	(0.288, 0.947)
	Urinary chromium	0.866	(0.630, 0.983)
10	Styrene	0.297	(0.000, 0.801)
	Urinary mandelic acid	0.267	(0.000, 0.789)
	Urinary phenylglyoxylic acid	0.083	(0.000, 0.705)
11	Carbon disulfide	0.196	(0.000, 0.866)
	Urinary TTCA	0.627	(0.000, 0.952)
12	Pyrene	0.512	(0.021, 0.897)
	Urinary 1-hydroxypyrene	0.497	(0.007, 0.892)
13	Styrene	0.677	(0.000, 0.946)
	Blood styrene	0.339	(0.000, 0.868)
	Urinary mandelic acid	0.474	(0.000, 0.903)
14	Inorganic mercury	0.397	(0.152, 0.788)
	Blood mercury	0.860	(0.693, 0.969)

[a]ICC calculated as $\hat{\sigma}_B^2/(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)$.
[b]Lower and upper confidence limits were negative.

of the air-biomarker and multiple biomarker data sets are summarized in Table III, which provides a profile of the two databases on the basis of type of industry and agent, residence time of the biomarkers, duration of monitoring, and sample size.

For the groups of workers on whom both personal and biological monitoring had been conducted, 15 of the 35 data sets (43%) were characterized by greater levels of between-worker variability than within-worker variability when the point estimates of the variances were compared. The point and interval estimates of the ICCs (based upon the model with a CS covariance structure) appear in Table IV. For these data sets, the estimated ICC ranged from 0 to 0.993 with a median value of 0.474. In stratifying the data by type of exposure measure, the ICC median value dropped from 0.627 for the biological measures (n = 19 data sets) to 0.263 for the airborne measures (n = 16). On a group by group basis, the estimated ICC values were larger for the biological monitoring data as compared to the personal sampling data for 8 out of the 14 groups (57%).

Table V summarizes the results for the 18 groups of workers whose exposures were evaluated with two or more biomarkers (42 sets of data). Information about the half-life of the measured compound (short = less than 7 days,

TABLE V. Point and Interval Estimates of the Intra-Class Correlation Coefficients (ICCs) for Workers' Exposures Assessed in Parallel by Multiple Biological Measures of Exposure (Based Upon the Model With Compound Symmetry)

Group	Type	Half-life[a]	ICC[b]	95% CI
15	Urinary inorganic arsenic	Short	0.834	(0.515, 0.952)
	Urinary methylarsonic acid	Short	0.481	(0.000, 0.825)
	Urinary dimethylarsinic acid	Short	0.417	(0.000, 0.798)
	Urinary trimethylarsenic compounds	Short	0.792	(0.419, 0.939)
16	Alveolar air toluene	Short	0.423	(0.041, 0.782)
	Blood toluene	Short	0.772	(0.495, 0.930)
	Urinary hippuric acid	Short	0.542	(0.167, 0.839)
17	Urinary cadmium	Long	0.162	(0.000, 0.636)
	Blood cadmium	Long	0.389	(0.000, 0.761)
18	Urinary tetrachlorophenol	Short	0.570	(0.396, 0.719)
	Urinary pentachlorophenol	Intermediate	0.509	(0.325, 0.674)
19	Urinary mandelic acid + phenylglyoxylic acid	Short	0.000	(0.000, 0.690)
	Urinary mandelic acid + phenylglyoxylic acid	Short	0.000	(0.000, 0.671)
20	Urinary chromium	Intermediate	0.723	(0.515, 0.892)
	SCEs/cell (mean)		0.712	(0.501, 0.887)
21	Plasma 2,4-TDA	Intermediate	0.942	(0.698, 0.992)
	Plasma 2,6-TDA	Intermediate	0.998	(0.987, 1.000)
22	Blood arsenic	Short	0.000	(0.000, 0.132)
	Urinary arsenic	Short	0.000	(0.000, 0.485)
23	Urinary lead	Intermediate	0.168	(0.000, 0.749)
	Urinary ALA	Long	0.457	(0.117, 0.894)
24	Blood lead	Intermediate	0.456	(0.087, 0.897)
	Urinary lead	Intermediate	0.637	(0.263, 0.943)
	Urinary ALA	Long	0.579	(0.199, 0.930)
25	Blood styrene	Short	0.342	(0.000, 0.764)
	Urinary mandelic acid	Short	0.268	(0.000, 0.728)
	Urinary phenylglyoxylic acid	Short	0.000	(0.000, 0.459)
26	Urinary inorganic As + MMA + DMA	Short	0.937	(0.615, 0.993)
	Urinary inorganic As + MMA + DMA	Short	0.841	(0.220, 0.982)
27	Blood styrene	Short	0.752	(0.159, 0.951)
	Urinary mandelic acid	Short	0.380	(0.000, 0.854)
	06-styrene-guanine lymphocyte adducts	Long	0.678	(0.008, 0.935)
28	Urinary nickel	Short	0.686	(0.374, 0.913)
	Urinary nickel	Short	0.610	(0.276, 0.886)
29	Blood mercury	Intermediate	0.330	(0.000, 0.793)
	Urinary mercury	Long	0.591	(0.000, 0.888)
30	Plasma cholinesterase	Intermediate	0.896	(0.820, 0.949)
	RBC cholinesterase	Intermediate	0.508	(0.315, 0.710)
31	Plasma cholinesterase	Intermediate	0.908	(0.802, 0.970)
	RBC cholinesterase	Intermediate	0.399	(0.143, 0.722)
32	Plasma cholinesterase	Intermediate	0.938	(0.872, 0.977)
	RBC cholinesterase	Intermediate	0.327	(0.090, 0.634)

[a]Short = less than 7 days, intermediate = 7 days to 4 weeks, long = greater than 4 weeks.
[b]ICC calculated as $\hat{\sigma}_B^2/(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)$.

intermediate = 7 days to 4 weeks, and long = greater than 4 weeks) is also given. Overall, 55% (n = 23) of the data sets were characterized by greater levels of between-worker variability than within-worker variability. Estimates of the ICCs ranged from 0 to 0.998 with a median value of 0.526. On the basis of residence time in the body, only five groups (groups 18, 23, 24, 27, and 29) were measured with biomarkers of different half-lives (''short,'' ''intermediate''

or ''long''). For group 18, the estimated ICC values for the ''short-lived'' biomarker as compared to the ''intermediate-lived'' biomarker were nearly the same (0.570 and 0.509 for urinary tetrachlorophenol and pentachlorophenol, respectively). Likewise, results on group 24 for urinary lead (''intermediate'') and urinary ALA (''long'') were similar (0.637 and 0.579). Biomarkers with longer residence times were characterized by larger ICC values for group 23 (0.168 and 0.457 for urinary lead (''intermediate'') and urinary ALA (''long''), respectively), group 24 (0.456 and 0.637 for blood lead (''intermediate'') and urinary lead (''long''), respectively) and group 29 (0.330 and 0.591 for blood mercury (''intermediate'') and urinary mercury (''long''), respectively).

In evaluating the results from the random-effects model with an exponential correlation error structure, significant serial correlation ($P < 0.05$) was detected in data collected on tobacco farmers (groups 30, 31, and 32; both measures), chloralkali processing plant workers (group 14; blood mercury), sawmill workers (group 18; both measures), forestry workers (group 22; blood arsenic) and alkyl lead manufacturing workers (group 24; urinary lead). For these groups, the estimated values of the ICCs under the one-way random effects model with either a CS or exponential correlation error structure, computed on the basis of the number of measurements collected on each worker (Equations 4 and 5, respectively), are shown in Table VI. Generally, the ICCs appear to be over-estimated when serial correlation was present but ignored in the analyses.

TABLE VI. Estimated Intra-Class Correlation Coefficients Calculated Under Models With an Exponential Error Structure [ICC(EXP)] and Compound Symmetric (CS) Structures [ICC(CS)] for Groups With Serially Correlated Measurements ($P < 0.05$)

Group	Type	N[a]	$\hat{\rho}^{1\,day}$[b]	ICC (EXP)[c]	ICC (CS)[d]
14	Blood mercury	8	0.842	0.967	0.980
18	Urinary tetrachlorophenol	4	0.991	0.782	0.844
	Urinary pentachlorophenol	4	0.983	0.788	0.807
22	Blood arsenic	8	0.916	0	0
24	Urinary lead	5	0.823	0.860	0.898
30	Plasma cholinesterase	5	0.845	0.971	0.977
	RBC cholinesterase	5	0.966	0.384	0.835
31	Plasma cholinesterase	5	0.992	0.624	0.980
	RBC cholinesterase	5	0.902	0.653	0.768
32	Plasma cholinesterase	5	0.868	0.982	0.986
	RBC cholinesterase	5	0.947	0.180	0.707

[a]Number of repeated measurements collected on each worker.
[b]Estimated correlation coefficient for a 1-day interval between measurements.
[c]ICC calculated as $\dfrac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \frac{\hat{\sigma}_W^2}{n}\left(1 + \frac{2}{n}\sum_{j=2}^{n}\sum_{j'=1}^{j-1}\hat{\rho}^{\tau_{jj'}}\right)}$.
[d]ICC calculated as $\hat{\sigma}_B^2/(\hat{\sigma}_B^2 + \hat{\sigma}_W^2/n)$.

DISCUSSION

In this study, multiple measures of exposure collected in tandem from the same workers were examined in view of the fact that each measure serves as a surrogate of a worker's true exposure and that quantifying the degree of measurement error inherent in each is an important (albeit not the sole) criterion for evaluating the utility of an exposure measure. In evaluating different measures of exposure, comparisons may be limited because of differences in exposure pathways, underlying kinetics, or mechanisms of action as would be the case, for example, in comparing urinary levels of inorganic arsenic to methylated arsenic. Notwithstanding this limitation, we found that airborne measurements typically were characterized by greater levels of variation from one sampling period to another relative to the variation between workers as compared to concurrently collected biological measurements. In general, this finding suggests that there is more measurement error in airborne measures than in biological measures with a greater likelihood of biasing effects in epidemiologic studies. Our results are consistent with findings reported in a recent study [Lin et al., 2005] that compared air sampling and biological monitoring data, but without the strict criteria applied in the current investigation that required that comparisons be restricted to concurrent sampling on identical sets of individuals over common sampling intervals.

While biological measurements appear to be preferable to airborne measurements in many cases, collecting additional airborne measurements may be more feasible and cost-effective to minimize the bias in exposure-response associations in the presence of exposure measurement error. Nonetheless, we may be willing to tolerate more error in a biomarker than in an air sample because it may be deemed a more relevant measure of exposure due to the kinetic processes involved that could, for example, result in quite different doses at a target site even in the face of similar external exposure profiles among individuals. Moving away from the question of selecting one exposure measure rather than another, cogent arguments have been made that support the collection of both airborne and biological monitoring data to fully explore the relationships between exposure and dose, as well as the roles that inhalation and dermal pathways may play in influencing body burden [Lin et al., 2005].

Although limited in the number of comparisons that could be made on a group-by-group basis between different biological measures of exposure on the basis of their residence time, biomarkers with longer half-lives tended to be characterized by less measurement error, as would be expected given the smoothing of variability of a contaminant in air levels as compared to biological levels [Rappaport, 1985]. Nonetheless, our results also suggest that multiple measurements may be required to reliably estimate worker exposure even for biomarkers with half-lives that extend over

relatively longer periods. For example, assume that investigators wished to limit the attenuation of a simple linear regression coefficient (based on the simple measurement error model applied in this study) to 10% for workers exposed to mercury in a chemical manufacturing plant (group 29). Here, our results suggest that 6 or 18 measurements would be required on each worker depending on whether urine or blood samples, respectively, were collected. Thus, even for biomarkers with relatively long half-lives in the body, it may become necessary to rely upon longitudinal data to estimate exposure.

An underlying assumption of the measurement error model applied in this study is that the mean exposure level is the relevant exposure metric, whereas such a model would not be applicable if peaks in exposure levels were of primary interest [Checkoway and Rice, 1992]. In considering how long the monitoring period should be to estimate workers' mean levels, decisions will be dictated by the exposure-response association under investigation. Since much of the data that were analyzed in this study were collected over relatively short intervals, any statements that are made regarding the magnitude of bias in exposure-response associations due to exposure measurement error are likely constrained by similarly short time frames. On the other hand, if interest was focused on estimating worker exposure over longer periods, then it would become necessary (assuming that sufficient data were collected) to model both the random and systematic sources of variation in exposure due to time effects. This becomes critical when evaluating the magnitude of exposure measurement error since improperly specifying the mean structure can result in biased estimates of the variance components [Symanski et al., 1996]. In addition, if measurements closer together in time are more highly correlated than those spaced farther apart, then variance components and functions thereof estimated from models that ignore serial correlation may be biased as shown previously [Symanski and Greeson, 2002] and in this investigation. While we found little evidence of serially dependent measurements ($<15\%$ of the data), the residence time of a biomarker should be factored into decisions regarding the longitudinal design of a study and the analytic approaches used for modeling the correlation structure among multiple measurements collected on the same individual.

Even though our results suggest that biomarkers of exposure in general are characterized by less error than airborne measures and therefore would introduce less bias in a health effects study, the data upon which these findings are based relied upon relatively small sample sizes. Indeed, the interval estimates for the ICC were wide in many cases, indicating that more workers and additional samples would be required to obtain more precise estimates. While the purpose of many of the studies from which the exposure data were abstracted was not necessarily to quantify the intra- and inter-individual sources of variation in exposure, issues related to small sample sizes plague efforts to accurately assess exposure in general and to quantify sources of variation in particular [Symanski et al., 2006]. As such, we not only concur with recommendations made previously in the literature that pilot studies be carried out to evaluate the relative magnitude of the within- and between-person sources of variation, but also further urge investigators to collect sufficient data to reliably estimate individual workers' mean exposure levels and thereby allow for meaningful inferences that depend upon these estimates in epidemiologic studies.

Finally, the groups of workers represented in this study were diverse in terms of the contaminants to which they were exposed, the type of workplaces monitored, and the sampling strategies employed to determine when, how often and at what interval to collect measurements. Even in those instances where the same contaminants were measured for two or more groups, differences make it difficult to generalize on the basis of any pair of exposure measures and caution is warranted in using results from one study to correct for measurement error in another [Carroll and Galindo, 1998]. Such differences may manifest as a result of factors influencing exposure in the workplace or other determinants of body burdens of contaminants, including, but not limited to, the timing of sample collection (for short-lived biomarkers especially), dermal absorption, and non-occupational sources of exposure. Given the difficulties of transporting measurement error results across study populations, the utility of evaluating exposure variability is underscored once more in the planning and conduct of occupational epidemiologic studies [Loomis and Kromhout, 2004].

In summary, our article highlights the importance of quantifying the between- and within-worker sources of variation in exposure, which in turn provides useful information about the degree of measurement error that may be present in an individual-based exposure assessment. Our findings indicate that there was more measurement error in biomarkers with shorter rather than longer half-lives and in airborne measures than in biological measures among those groups studied, although both types of monitoring data provided examples of exposure measures fraught with error. Our results also indicate substantial imprecision in the estimates of exposure measurement error, suggesting that greater attention needs to be given in the future to studies that collect sufficient data to allow for a better characterization of the attenuating effects of an error-prone exposure measure.

REFERENCES

Armstrong BG. 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med 55:651–656.

Carroll RJ, Galindo CD. 1998. Measurement error, biases, and the validation of complex models for blood lead levels in children. Environ Health Perspect 106 (Suppl 6):1535–1539.

Checkoway H, Rice CH. 1992. Time-weighted averages, peaks, and other indices of exposure in Occupational Epidemiology. Am J Ind Med 21:25–33.

Fuller WA. 1987. Measurement error models. New York: Wiley.

Liljelind I, Rappaport S, Eriksson K, Andersson J, Bergadahl IA, Sunesson A-L, Järvholm B. 2003. Exposure assessment of monoterpenes and styrene: A comparison of air sampling and biomonitoring. Occup Environ Med 60:599–603.

Lin YS, Kupper LL, Rappaport SM. 2005. Air samples versus biomarkers for epidemiology. Occup Environ Med 62:750–760.

Loomis D, Kromhout H. 2004. Exposure variability: Concepts and applications in occupational epidemiology. Am J Ind Med 45:113–122.

Rappaport SM. 1985. Smoothing of exposure variability at the receptor: Implications for health standards. Ann Occup Hyg 29:201–214.

Rappaport SM, Symanski E, Yager JW, Kupper LL. 1995. The relationship between environmental monitoring and biological markers in exposure assessment. Environ Health Perspect 103 (Suppl 3):49–54.

Searle SR, Casella G, McCulloch CE. 1992. Variance components. New York: Wiley. p 65–66.

Symanski E, Bergamaschi E, Mutti A. 2001. Inter- and intra-individual sources of variation in levels of urinary styrene metabolites. Int Arch Environ Occup Health 74:336–344.

Symanski E, Sällsten G, Barregård L. 2000. Variability in airborne and biological measures of exposure to mercury in the chloralkali industry: Implications for epidemiologic studies. Environ Health Perspect 108: 569–573.

Symanski E, Greeson NMH. 2002. Assessment of variability in biomonitoring data using a large database of biological measures of exposure. Am Ind Hyg Assoc J 63:390–401.

Symanski E., Kupper LL, Kromhout H, Rappaport SM. 1996. An investigation of systematic changes in occupational exposure. Am Ind Hyg Assoc J 57:724–735.

Symanski E, Maberti SI, Chan W. 2006. A meta-analytic approach for characterizing the within- and between-worker sources of variation in exposure to workplace contaminants. Ann Occup Hyg 50:343–357.

Teschke K, Spierings J, Marion SA, Demers PA, Davis HW, Kennedy SM. 2004. Reducing attenuation in exposure-response relationships by exposure modeling and grouping: The relationship between wood dust exposure and lung function. Am J Ind Med 46:663–667.

Thomas D, Stram D, Dwyer J. 1993. Exposure measurement error: Influence on exposure disease relationships and methods for correction. Annu Rev Public Health 14:69–93.

Tjoe E, Höhr D, Borm P, Burstyn I, Spierings J, Steffens F, Lumens M, Spee T, Heederik K. 2004. Variability in quartz exposure in the construction industry: Implications for assessing exposure-response relations. J Occup Environ Hyg 1:191–198.

van Tongeren MK, Gardiner K, Calvert I, Kromhout H, Harrington JM. 1997. Efficiency of different grouping schemes for dust exposure in the European carbon black respiratory morbidity study. Occup Environ Med 54:714–719.

United Nations. 1990. International Standard Industrial Classification of All Economic Activities (Statistical papers series M, No 4, Rev*). New York: United Nations.