Journal of Exposure Science and Environmental Epidemiology (2017) 27, 582–590
www.nature.com/jes

OPEN

ORIGINAL ARTICLE

Variation in urinary spot sample, 24 h samples, and longer-term average urinary concentrations of short-lived environmental chemicals: implications for exposure assessment and reverse dosimetry

Lesa L. Aylward[1], Sean M. Hays[2] and Angelika Zidek[3]

Population biomonitoring data sets such as the Canadian Health Measures Survey (CHMS) and the United States National Health and Nutrition Examination Survey (NHANES) collect and analyze spot urine samples for analysis for biomarkers of exposure to non-persistent chemicals. Estimation of population intakes using such data sets in a risk-assessment context requires consideration of intra- and inter-individual variability to understand the relationship between variation in the biomarker concentrations and variation in the underlying daily and longer-term intakes. Two intensive data sets with a total of 16 individuals with collection and measurement of serial urine voids over multiple days were used to examine these relationships using methyl paraben, triclosan, bisphenol A (BPA), monoethyl phthalate (MEP), and mono-2-ethylhexyl hydroxyl phthalate (MEHHP) as example compounds. Composited 24 h voids were constructed mathematically from the individual collected voids, and concentrations for each 24 h period and average multiday concentrations were calculated for each individual in the data sets. Geometric mean and 95th percentiles were compared to assess the relationship between distributions in spot sample concentrations and the 24 h and multiday collection averages. In these data sets, spot sample concentrations at the 95th percentile were similar to or slightly higher than the 95th percentile of the distribution of all 24 h composite void concentrations, but tended to overestimate the maximum of the multiday concentration averages for most analytes (usually by less than a factor of 2). These observations can assist in the interpretation of population distributions of spot samples for frequently detected analytes with relatively short elimination half-lives.

Journal of Exposure Science and Environmental Epidemiology (2017) **27,** 582–590; doi:10.1038/jes.2016.54; published online 5 October 2016

Keywords: biomonitoring; epidemiology; personal exposure

INTRODUCTION

The Canadian Health Measures Survey (CHMS) collects biological samples (blood, urine) from a representative sample of the Canadian population, and the National Health and Nutrition Examination Survey (NHANES) does the same for the US population. Spot urine samples are generally collected in such surveys rather than 24 h urine samples for logistical purposes. Many of the compounds that are measured in urine samples are relatively short-lived in the body (short biological half-lives) and urinary concentrations are expected to vary both within and between individuals, both within and across days.[1,2]

Urinary biomonitoring data are increasingly being used in a risk-assessment context to estimate daily intakes of various compounds through the application of reverse dosimetry approaches.[3–9] The biomonitoring data are valuable because such data reflect exposure from multiple product uses, exposure to multiple environmental media, and through dietary exposure. Reverse dosimetry relies on estimates of daily urinary flow or creatinine excretion to extrapolate from the observed concentration in urine to an estimated daily excretion rate of the analyte. The calculated daily excretion rate of the analyte is then adjusted using estimates of the fraction of ingested compound that is excreted in urine, as the analyte, which may be a parent compound or one or more metabolites, to back-calculate daily intake amounts of the parent compound. These daily intakes can then be divided by assumptions regarding a standard or average bodyweight to estimate intakes in mg/kg-day of parent compound, a dose metric that can be compared with previously derived tolerable intakes or to a toxicological point of departure.

In most exposure and risk assessment contexts, the average concentration of a compound in urine in an individual is more interpretable in terms of daily intake rates and of greater interest than the extremes of short-term fluctuations in concentration that may be captured in spot sampling. That is, in general, risk assessments are based on assessments of exposure-response using chronic exposure studies and tolerable intakes are assessed on an average basis over time. However, as discussed above, for logistical reasons, studies usually collect only spot urine samples. Thus, interpretation of the distributions of population-based urinary biomonitoring data for such compounds would be enhanced by understanding the relationships between the distribution of measured spot sample concentrations and the

[1]Summit Toxicology, LLP, Falls Church, Virginia, USA; [2]Summit Toxicology, LLP, Bozeman, Montana, USA and [3]Health Canada, Ottawa, Ontario, Canada. Correspondence: Dr. Lesa L. Aylward, Summit Toxicology, LLP, 6343 Carolyn Drive, Falls Church, VA 22044 USA.
E-mail: laylward@summittoxicology.com
Received 14 September 2015; accepted 21 August 2016; published online 5 October 2016

distribution of underlying longer-term average biomarker concentrations.

This report provides an analysis of the relationship between spot urine concentrations and longer-term averages (e.g., 24 h average and longer where adequate data exist). The following specific questions are addressed:

1. How does the distribution of spot samples, which are collected at times that are convenient for the participant or investigator, compare with the distribution of composite 24 h sample concentrations? In particular, because spot sample collections in populations are often examined in the context of risk assessments for chemical exposures in the population, how do upper percentiles (e.g., 95th percentile) in spot sample distributions compare with the corresponding percentile of longer-term average concentrations (e.g., 24 h sample concentrations or multiday average concentrations)?
2. How much variation in composite 24 h sample concentrations is there within and between individuals, and how does that compare with spot sample variation and with longer-term average variation?
3. How accurate are reverse dosimetry estimates based on upper-end metrics from the spot sample distributions for characterizing the upper end of the range of population intake rates?

METHODS

To address these questions, we provide analyses of two data sets in which multiple individuals collected every urine void for several days. The analyses included both parametric and nonparametric evaluations of these data sets to examine patterns both within and across the two data sets.

Data Sets

The data sets were from a study at the Centers for Disease Control and Prevention (CDC) and from a study conducted by a group of researchers from Europe and the United States under a research project funded by the European Chemical Industry Association (CEFIC). Each data set is comprised of serial spot urine samples collected over the course of several days in a group of eight individuals.

CDC serial sampling data set. The U.S. Centers for Disease Control and Prevention (CDC) recruited eight adult volunteers, four male and four female, ages 25 to 58, to collect every urine void for 7 continuous days during 2005.[2,10,11] Of the analytes analyzed by this group, bisphenol A (BPA), monoethyl phthalate (MEP — metabolite of diethyl phthalate exposures) and mono(2–ethyl–5–hydroxyhexyl) phthalate (MEHHP — metabolite of di(2-ethylhexyl) phthalate exposures) were included in this analysis as these are ingredients in consumer products or trace contaminants in food with high rates of detection in the data set. The raw data from this study were obtained from CDC. Of the total 56 days (8 participants × 7 collection days) possible, only 44 of the person-days included complete data (based on self report of all urine voids collected and measured). Our analysis was limited to these 44 person-days. The total number of days included in the multiday averages presented for each individual varied on the basis of the number of complete days of urine collection (3 with 7 days, 2 with 6 days, 2 with 5 days, 1 with 1 day). All of the volunteers resided in the United States (US) at the time of the study. As a result, their exposures are dictated by personal habits and product formulations consistent with the US market.

CEFIC serial sampling data set. A multi-center collaborative research project in Europe was conducted in which eight volunteers, four male and four female, ages 31 to 66, collected all urine voids over 6 continuous days during 2011 with a focus on analytes expected to be present in personal care products.[12,13] Of the analytes included in this study, data for triclosan, BPA and methyl paraben were included in the current analysis because these analytes had the highest detection rates, allowing for a characterization of the full range of concentrations present in urine samples. This study involved a protocol in which the volunteers were instructed to abstain from using specific consumer products that contained the target analytes during 2 days in the middle of the collection period.[12,13] The urinary spot sampling data for these 2 days were omitted from our analyses for triclosan and methyl paraben as the specific abstention days resulted in non-typical exposure patterns, resulting in 4 days of urine spot sample data from days with typical product use patterns remaining in the data set for this analysis. As BPA was not an ingredient in any of the personal care products used by the participants, and exposures to BPA were therefore not related to specific personal care product use but instead reflected exposure from other pathways, all 6 days of collected urine spot sample data for BPA were included in our analyses here. All volunteers resided in Belgium at the time of the study. As a result, their exposures are dictated by personal habits and product formulations in Belgium.

Analytes

The data for BPA were available from both sampling data sets. Methyl paraben and triclosan were available in the CEFIC data set, and MEP and MEHHP were included in the CDC data set. Likely sources, routes of exposure, and urinary elimination half-life are provided in Table 1 for each of the analytes examined in this study. These characteristics have some influence on the degree of inter- and intra-individual variability, both because of the influence on exposure pattern (for example, once per day in a personal care product in the morning *vs* several times per day for a chemical widely present as a trace component in foods) and because of the influence of the relationship between elimination half-life and exposure intervals.[1] Within-individual variation in sample concentrations will increase as the ratio of the half-life of elimination to the interval between exposures decreases.

Statistical Analyses

The data from these two studies allow comparison of spot urine concentrations to 24 h composite average and longer-term averages. The 24 h composite averages were calculated as total mass of analyte excreted divided by total urine volume over each 24 h period. Likewise, the multiday averages (4, 6, or 7 days) were calculated as total mass excreted over the corresponding days divided by the total volume of urine excreted over the same days. All the statistical analyses were conducted using Stata (version 12, www.stata.com).

Summary statistics including arithmetic and geometric means and standard deviations and key percentiles for spot, 24 h composite, and multiday average concentrations were calculated. Sampling data were plotted, inspected, and found to be generally lognormally distributed. Intraclass correlation coefficients (ICCs) for both spot and 24 h composite sample concentrations were also calculated. For each type of sample (spot or 24 h composite), a mixed effects model was implemented on log-

Table 1. Characteristics of the analytes included in this analysis.

Compound	*Example sources of exposure*	*Route of exposure*	*Half-life of elimination (h)*
MEP	Consumer products	Dermal, inhalation	No formal estimates available; likely < 6[a]
Methyl paraben	Consumer products	Dermal	6–8[b]
Triclosan	Toothpaste, soaps, deodorants, sanitizers, cleaners	Oral, dermal	11 (ref. 20)
BPA	Food contamination	Oral	4–6 (ref. 15)
MEHHP	Food contamination	Oral	10 (ref. 18)

[a]Based on observation of individual urinary concentration *vs* time profiles in Preau et al.[2] [b]Terminal excretion half-life observed in three human volunteers administered a controlled dose of deuterated methyl paraben. Personal communication, Holger Koch, 20 February 2015.

transformed concentrations to assess the within- and between-subject variance. The ICC values were calculated as the ratio of the between-person variance to the total variance (within+between).

The ICC values vary between 0 and 1. Higher values of the ICC (greater than 0.5) indicate that variance between individuals is greater than variance within individuals. As the ICC increases, the reliability of a single spot sample for characterizing or categorizing the longer-term concentrations in that individual relative to other individuals increases. Conversely, for chemicals with low ICC values, within-person variation is large compared with between-person variation, and individual spot samples are not likely to provide a reliable characterization of an individual's longer-term relative exposure level. The ICC values for 24 h composite samples likewise provide an indication of the relative consistency from day-to -day of the 24 h average urinary concentrations in an individual.

To examine the relationship between spot, 24 h composite, and multiday composite concentrations, we compared selected statistics from the distributions. We compared the GMs of the spot and 24 h composite sample distributions to evaluate how well the spot sample GM predicts the 24 h GM. We also compared the 95th percentile (P95) of the spot sample distributions to the P95 of the 24 h composite sample distribution as well as to the maximum of the multiday average concentrations observed in the eight individuals for a given analyte. We also assessed the impact of considering only spot samples collected during the typical time periods included in the NHANES program (roughly, between 9:00 and 21:00, based on the morning, afternoon, and evening sessions for NHANES participants).

Finally, we plotted the participant-by-participant spot, 24 h composite, and multiday average concentrations for each analyte to allow visual assessment of the within- and between-individual variations in these metrics as well as to allow visual assessment of the relationship between the distributions for individuals and the distributions for the populations as a whole.

Reverse Dosimetry Example

As discussed above, with the availability of population-representative urinary biomonitoring data for a number of substances of interest, regulatory agencies have begun to rely on such data to estimate population exposure rates through the application of reverse dosimetry approaches.[4,5,9] The data sets collected here present an opportunity to truth-test such efforts because the total mass of excreted analytes can be calculated for each individual in the data sets over the course of the multiple days included (sum of the product of void volume and analyte concentration from each void).

For this comparison, we take the 95th percentile (P95) from the spot sample distribution for each of the analytes in the respective data sets. We then calculated the estimated daily intake (DI) in μg/day associated with that percentile as follows:

$$\mathrm{DI} = \frac{C_{95} \times V_{24}}{\mathrm{FUE}} \quad (1)$$

where C_{95} is the measured concentration at the P95 from the spot sample distribution, V_{24} is an estimate of average 24 h urine volume for adults. Daily urine flow varies widely within and between individuals. For this estimate, we assumed 1.7 liter/day, which is the approximate average daily void volume for adults measured in a large study.[14] FUE is the mass urinary excretion fraction for the analyte (mass of analyte excreted in urine/mass of parent compound ingested). FUE values used here were 1 for BPA;[15] 0.17 for methyl paraben;[16] 0.185 for MEHHP;[17,18] 0.54 for triclosan;[19,20] and 0.61 for MEP.[3]

A similar calculation can be conducted on the basis of creatinine-adjusted concentrations by assuming a daily creatinine excretion rate in place of a daily urine volume. Creatinine excretion rate can be estimated on a person-specific basis based on gender, height, weight, and age using the equations of Mage et al.,[8] or a standard central tendency estimate can be used. For this illustrative effort, a standard value of 1.3 g creatinine excreted per day for adults was used;[8] however, other estimates are available.

We calculated for each subject-day the total estimated mass intake of the parent compound by summing the excreted analyte mass (sum of the product of the volume of each void times the measured concentration) divided by the appropriate FUE as described above. We compared the estimated DI for each analyte based on the spot sample P95 to the P95 of the distribution of the calculated daily intakes from all subject-days and to the maximum of the multiday average of the daily intakes for all eight subjects in the respective study.

RESULTS

The summary statistics for the distributions of spot, 24 h composite, and multiday average concentrations for each of the included analytes from each study are presented in Table 2. Figures 1 and 2 present the distribution of measured urinary concentrations for spot samples, 24 h composite samples, and the multiday averages for each individual in the two data sets for the selected analytes.

Variation in 24 h Averages Within and Between Individuals

Table 2 presents the summary statistics for 24 h composite samples over all the participants in each study for each analyte. Because of the limited number of 24 h composites for each individual (four to seven), no statistics were calculated for the person-by-person composites. However, visual inspection of Figures 1 and 2 allow some characterization of the within-individual variation. In addition, the ICC values in Table 3 for the 24 h composites provide a quantitative evaluation of the relative between-individual variation in those composite concentrations.

For most analytes (MEP, BPA, triclosan, methyl paraben), the variation in 24 h averages within an individual spans an order of magnitude or less over the multiple days examined here. MEHHP was the only analyte for which the variation in 24 h averages spans greater than an order of magnitude within several individuals, spanning almost two orders of magnitude for some individuals, which is approximately the same degree of variation in 24 h averages across the entire study population. This suggests highly heterogeneous exposure patterns to DEHP.

Intraclass Correlation Coefficients

Intraclass correlation coefficients are presented in Table 3. Intraclass correlation coefficients range from very low (ICCs less than 0.2 for BPA) to very high (greater than 0.9 for MEP and triclosan). The ICCs for 24 h composite samples are consistently higher than those for spot samples, reflecting the more stable nature of the 24 h composite samples. The ICCs indicate that single spot samples would be relatively unreliable for categorizing relative exposure level for an individual in a population for BPA, slightly more reliable for DEHP (based on MEHHP), moderately reliable for methyl paraben, and most reliable for MEP and triclosan. The compounds with the highest ICC are those compounds for which we would expect noticeable differences in exposure among individuals resulting from lifestyle/consumer product choices (e.g., methyl paraben, triclosan, MEP). Those compounds with low ICC are consistent with the compounds that we would expect all individuals to have more similar exposures across individuals and days resulting from general food contamination (BPA, DEHP). However, it is important to note that a high ICC value does not necessarily indicate that spot samples within an individual have similar concentrations over time. Spot samples within and between days for an individual may still vary widely, even in the presence of a high ICC coefficient (see, for example, Figures 1 and 2). The high ICC values for MEP and triclosan do not preclude within-individual variation in spot sample concentrations spanning more than two orders of magnitude for MEP and one order of magnitude for triclosan. ICCs calculated on the basis of creatinine-adjusted concentrations tended to be somewhat higher than those from unadjusted concentrations; however, this was not universally true (see Table 3).

The range of concentrations observed both within individuals and across the full data set narrows as you progress from shorter to longer time of collection (spot *vs* 24 h *vs* multiday averages; Figures 1 and 2, Table 2). The total reduction in range for the all-participant spot samples compared with the all-participant multi-day average (Figures 1 and 2) is relatively smaller for those

Table 2. Summary statistics for concentrations (ng/ml) in spot samples, 24 h composite samples, and multiday composites for analytes from two serial urinary collection studies.

	N	% >LOD (LOD, ng/ml)	Mean	SD	GM	GSD	P5	P25	P50	P75	P90	P95	Min	Max
Methyl paraben, CEFIC														
Spot	229	100 (1)	168.1	427.6	55.2	4.4	4.9	17.5	61.3	151.0	317.0	432.0	1.3	4180.0
24-h composites	32		145.0	260.0	59.7	3.8	3.2	10.4	64.7	161.0	189.9	940.8	5.2	1226.1
4-day average	8		131.2	176.1									6.2	545.3
Triclosan, CEFIC														
Spot	229	80 (1)	570.5	715.5	104.4	17.7	0.7	56.5	314.0	767.0	1680.0	2050.0	0.7	3990.0
24-h composites	32		487.3	492.4	102.5	18.6	0.7	30.4	364.8	667.2	1254.6	1626.6	0.7	1824.2
4-day average	8		473.5	472.2									0.7	1426.7
BPA, CEFIC														
Spot	349	72 (0.5)	2.1	4.0	1.2	2.7	0.4	0.4	1.2	2.4	4.4	5.8	0.0	64.0
24-h composites	48		1.7	1.4	1.4	2.0	0.4	0.9	1.2	2.2	3.4	4.7	0.4	7.0
6-day average	8		1.6	0.6									0.7	2.6
BPA, CDC														
Spot	328	91 (0.4)	2.9	4.0	1.7	2.8	0.3	0.9	1.7	3.6	6.6	9.8	0.3	50.0
24-h composites	44		2.8	2.1	2.2	1.9	0.8	1.5	2.2	3.2	5.9	6.2	0.7	11.8
7-day average	8		2.5	1.3									1.2	4.5
MEP, CDC														
Spot	328	100	256.9	654.2	59.3	5.6	4.7	16.5	48.6	218.7	611.0	1410.8	0.3	7809.2
24-h composites	44		274.8	343.3	100.1	4.8	10.5	26.4	68.5	442.3	818.2	894.4	7.9	1228.6
7-day average	8		250.3	324.5									12.5	881.1
MEHHP, CDC														
Spot	328	>95	85.9	188.7	22.9	5.5	1.5	7.9	22.3	72.5	210.5	370.6	0.2	1629.6
24-h composites	44		72.8	85.3	37.8	3.3	6.5	13.5	33.4	119.6	173.6	278.6	4.5	371.2
7-day average	8		59.1	51.1									7.1	132.3

Multiday average values from the CEFIC study for methyl paraben and triclosan cover 4 days with the use of typical products with these ingredients, and for BPA all 6 study days (no intervention regarding BPA exposure). Multiday average values for the CDC study cover all 7 days of the study.

chemicals with a high ICC (triclosan, methyl paraben — Table 3) compared with those with a low ICC (BPA, MEHHP).

Relationships Among Key Distribution Statistics, Spot *vs* Longer-Term Averages

Table 4 presents comparison of key statistics from the distributions of concentrations of analytes in spot samples, 24 h composite samples, and multiday averages. The GM of the spot sample distribution generally under-predicted the corresponding GM of the distribution of 24 h composite samples for the analytes included here, except triclosan, for which the GMs were concordant. The ratio of unadjusted spot to 24 h GMs ranged from 0.6 for MEP and MEHHP to 1.0 for triclosan; creatinine adjustment produced similar, but slightly better, agreement, with the ratios ranging from 0.7 for MEP and MEHHP to 1.1 for triclosan. For most of the included analytes, the P95 of the spot sample distribution over-predicts the P95 of the 24 h composite sample distribution by a factor of 1.2 to 1.6, although for methyl paraben the spot sample P95 under-predicted the estimated P95 of the 24 h composite sample distribution (Table 4). However, inspection of the cumulative distribution function for methyl paraben suggests that, given the uncertainties associated with the small sample sizes here, the distributions of the spot and 24 h composite samples are more or less indistinguishable at the upper percentiles (data not shown). The pattern in ratios of P95 spot:24 h composites were similar for creatinine-adjusted concentrations compared with unadjusted concentrations.

Finally, the P95s of the spot sample distributions were compared with the range of multiday average concentrations for the eight individuals (Table 4). For methyl paraben, the P95 of spot samples (unadjusted) slightly under-predicted the maximum multiday average unadjusted concentration among the participants. However, for all other analytes, the spot sample P95 is higher than the maximum of the multiday average concentrations in the underlying data (1.4- to 2.8-fold for unadjusted concentrations and 1.1- to 2.6-fold for creatinine-adjusted concentrations). The pattern was similar when the spot samples were restricted to those collected during the non-nighttime hours (9:00 to 21:00; Table 4).

Reverse Dosimetry Example

The results of the calculations comparing reverse dosimetry-derived estimates of daily intake to those calculated from the actual excreted masses of the study analytes are presented in Figure 4. The daily intakes calculated from reverse dosimetry using the P95 of the spot sample concentrations, either as measured or creatinine-corrected, are compared with two metrics from the range of daily intakes calculated from the excreted analyte masses: the P95 of the full set of subject-day daily intake estimates (e.g., four day-specific estimates for each of the eight individuals in the CEFIC data set for methyl paraben, for 32 individual estimates; panel a in Figure 4), or the maximum of the multiday average intakes (i.e., the highest of the multiday average daily intake rates for the eight individuals) (panel b in Figure 4).

Reverse dosimetry from the spot sample P95 produced dose rate estimates that were similar to, or up to 50% higher than, the P95 of the daily intake rates calculated on the basis of the mass of excretion of analytes on a day-by-day basis (Figure 4, panel a). Similar results were obtained for the volume-based and creatinine-based calculations. When compared with the maximum average daily intake over the multiple days included in the urine collection, the reverse dosimetry estimates over-predict the

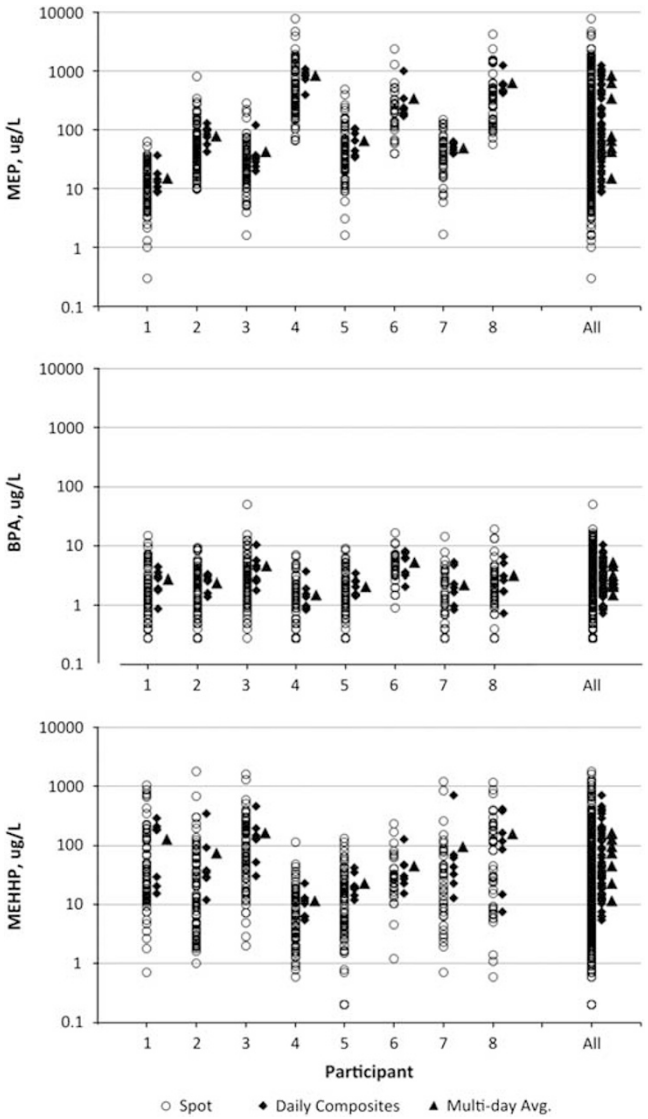

Figure 1. Plots of MEP, BPA, and MEHHP concentration by participant in the CDC cohort. For each participant, the spot, 24 h composite, and multiday average concentrations are plotted; the same metrics for all the participants are plotted in the last column.

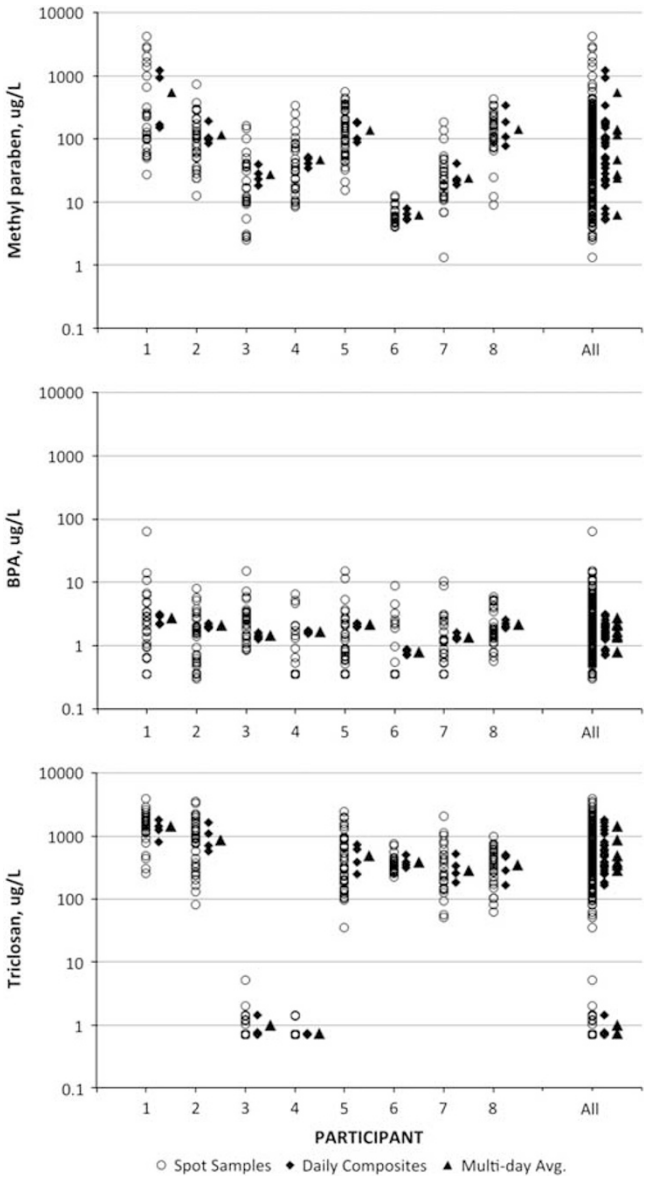

Figure 2. Plots of methyl paraben, BPA, and tricolsan concentration by participant from the CEFIC cohort. For each participant, the spot, 24 h composite, and multiday average concentrations are plotted; the same metrics for all participants are plotted in the last column.

Table 3. Intraclass correlation coefficients for spot samples collected over 4, 6, or 7 days (depending on study and analyte) and for 24 h composite samples in terms of unadjusted concentrations and creatinine-adjusted (cr-adjusted) concentrations.

Study	*Analyte*	*Log of unadjusted concentrations, ng/ml*		*Log of cr-adjusted concentrations, µg/g cr*	
		Spot samples	*24 h composites*	*Spot samples*	*24 h composites*
CDC	MEP	0.61	0.91	0.73	0.93
	MEHHP	0.23	0.39	0.21	0.33
	BPA	0.20	0.32	0.12	0.16
CEFIC	Methyl paraben	0.56	0.84	0.71	0.87
	Triclosan	0.93	0.98	0.96	0.99
	BPA	0.13	0.19	0.26	0.28

Concentrations were log-transformed for ICC calculations.

Table 4. Comparison of geometric mean (GMs) and geometric standard deviations (GSDs) and the observed 95th percentiles for the distributions of spot and 24 h composite samples in the two studies (see Table 2).

Study, compound	*GM (GSD)*		*Observed concentration @ 95th %ile*		*Maximum observed multiday average*	*Ratio, spot:24 h GMs*	*Ratio, spot:24 h 95th %iles*	*Ratio, spot 95th %ile to maximum multiday avg*	
	Spot	*24 h*	*All spots*	*Daytime spots*	*24 h*			*All spots*	*Daytime spots*
Unadj., ng/ml:									
CEFIC, MP	55.2 (4.4)	59.7 (3.8)	431.5	337	940.8	545.3	0.9	0.5	0.8	0.6
CEFIC, Triclosan	104.4 (17.7)	102.5 (18.6)	2050	1750	1626.6	1426.7	1.0	1.3	1.4	1.2
CEFIC, BPA	1.17 (2.7)	1.4 (2)	5.8	5.2	4.7	2.6	0.8	1.2	2.2	2.0
CDC, BPA	1.7 (2.8)	2.2 (1.9)	9.8	8.1	6.2	4.5	0.8	1.6	2.2	1.8
CDC, MEHHP	22.9 (5.5)	37.8 (3.3)	370.6	432	278.6	132.3	0.6	1.3	2.8	3.3
CDC, MEP	59.3 (5.6)	100.1 (4.8)	1410.8	1446	894.4	881.1	0.6	1.6	1.6	1.6
Cr-adj., µg/g cr:									
CEFIC, MP	70.3 (4.2)	73.1 (4.1)	630.8	580	744.6	411.1	1.0	0.8	1.5	1.4
CEFIC, Triclosan	133 (15.4)	125.5 (16.3)	1359.2	1446.5	1227.5	1075.7	1.1	1.1	1.3	1.3
CEFIC, BPA	1.6 (2.5)	1.9 (2.1)	7.2	6.6	7.8	3.8	0.8	0.9	1.9	1.7
CDC, BPA	2.6 (2.3)	2.9 (1.8)	11.4	11.8	6.7	6.2	0.9	1.7	1.8	1.9
CDC, MEHHP	35.2 (3.7)	49 (2.9)	482.8	594.9	304.8	182.7	0.7	1.6	2.6	3.3
CDC, MEP	91 (5.4)	132.1 (4.8)	2640.1	2932.9	2250.0	2496.6	0.7	1.2	1.1	1.2

Assessments are presented on the basis of unadjusted concentrations (ng/ml) and on the basis of cr-adjusted concentrations (µg/g cr). "Daytime" spots are those restricted to typical times of collection in NHANES, 9:00 to 21:00.

maximum value substantially for most analytes, by a factor of more than 3 in the case of MEHHP.

The degree of correspondence between the reverse dosimetry daily intake estimates and those based on measured mass excretion is directly dependent on the assumed values for daily urinary volume and daily creatinine excretion rate. In reverse dosimetry exercises, exposure is often estimated by multiplying observed spot sample concentration by an assumed daily urine volume. To be conservative, risk assessors may use an upper-end estimate of daily urine flow, which will result in an estimate of daily intake that may overestimate, but not likely underestimate, actual daily exposure. If a "conservative" value for urine flow is used, that is, an upper-end daily urine volume instead of an average estimate of 1.7 liter/day as used here, the resulting reverse dosimetry estimates would be correspondingly higher, and would result in further overestimates of both individual daily intakes and maximum multiday average intakes compared with the observed values for most of the analytes and data sets examined here.

DISCUSSION

The intensive sampling data presented here allow a detailed examination of the relationships between the distributions in urinary analyte concentrations in spot samples and metrics of greater interest in a risk-assessment context: daily average concentrations or averages over multiple days. Although the populations examined here are small, the distributions in spot sample concentrations are relatively similar to those observed in the NHANES survey for adults for the same analytes (Figure 3). Spot sample distributions for MEHHP and BPA in the CDC serial sampling data set are nearly identical to those observed in NHANES for adults. For MEP, the CDC serial sampling volunteers were generally exposed to somewhat lower levels than overall in the NHANES participants, but the shape of the distribution of spot sample concentrations is very similar. Thus, by intensively sampling a few individuals, the patterns of the distribution of various urinary biomonitoring metrics for these compounds provide some insight into the exposure patterns among these study populations, and by extension, into extrapolations made from surveys like NHANES or the Canadian Health Measures Survey (CHMS). Some key findings and observations from this analysis are presented below.

Relationship of Spot Sample Concentrations to Longer-Term Average Concentrations

Although the full distribution of spot sample concentrations is broader than the full distribution of 24 h average concentrations for each analyte, the upper ends of the distributions tend to converge (Tables 2 and 4; Figures 1 and 2). Again, the best correspondence is observed for creatinine-adjusted concentration distributions. Spot sample 95th percentiles tended to over-predict the range of multiday average concentrations by a greater degree, except for methyl paraben (Table 4). Conversely, for most analytes, GM concentrations from spot sample distributions tend to underestimate the GM concentration from the 24 h average distributions, with the exception of triclosan, for which the two metrics were approximately equal. This reflects the skewed nature of the underlying distribution of spot sample concentrations, which arises from the short elimination half-lives. The time-distribution of concentrations have relatively few high concentration spot collection which rapidly decline to low concentrations owing to rapid elimination rates. Unless exposure is frequent relative to the half-life of elimination, the distribution of spot samples will be dominated by samples with concentrations less than the 24 h arithmetic mean, which is most highly influenced by the high concentration periods. This likely contributes to the

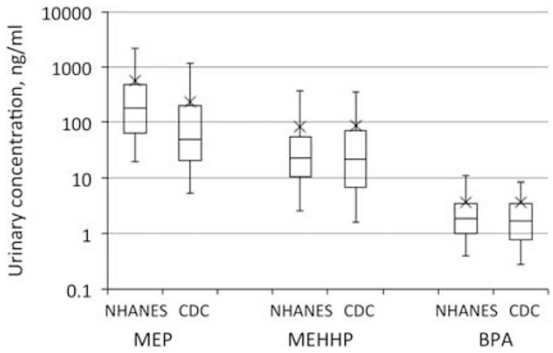

Figure 3. Comparison of the distributions of spot sample concentrations for MEP, MEHHP, and BPA in the CDC serial sampling data set ("CDC") to that from the NHANES survey in 2005–2006 for adults. The boxes represent the intraquartile range with the median as a horizontal line. Whiskers extend to the 5th and 95th percentiles, and the X indicates the arithmetic mean.

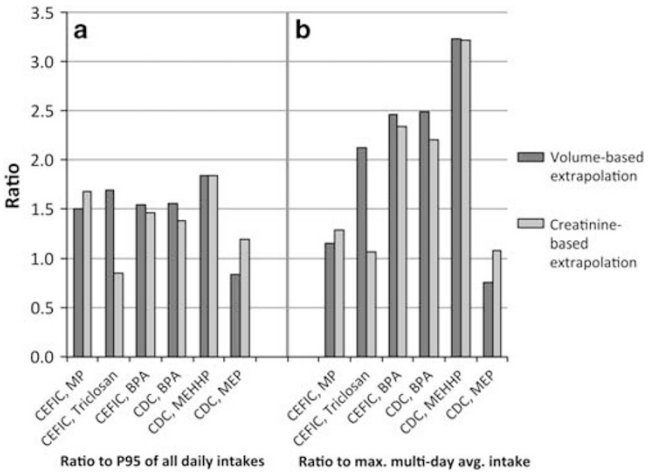

Figure 4. Comparison of estimates of daily intake derived through reverse dosimetry on the P95 of the distribution of spot sample concentrations (volume or creatinine-based) to the daily intake calculated from the mass of excreted analytes. The ratios are calculated compared with the P95 of the individual daily intakes across all individuals and days in the data set (a), or compared with the maximum multiday average daily intake for all the individuals in the data set (b). A ratio of 1 indicates an accurate correspondence between the reverse dosimetry-derived intake estimate and the two metrics derived from the mass excretion measured in the data sets, while higher ratios indicate an overestimation of actual intake.

closer concordance observed for triclosan, which has a somewhat longer elimination half-life than most of the other analytes included here.[20]

In a risk-assessment context, analytical results from spot samples collected in a population are often examined to estimate via reverse dosimetry the distribution of intake levels in the population. In this context, the focus is usually on estimating the upper end of the distribution of daily doses in the population. This analysis suggests that use of a 95th percentile from the distribution of population spot sample concentrations will provide a reasonable, if slightly conservative, estimate of the 95th percentile of the distribution of 24 h average concentrations. Intake rates estimated from these concentrations will provide a reasonable estimate of the upper end of daily intake rates in the population if average daily urine flow and creatinine excretion rates are used. This same metric will likely overestimate the upper percentiles of multiday average concentrations among individuals (and therefore longer-term average intake rates) for most transient analytes, with the degree of overestimation ranging from 10% up to a factor of 2. However, reliance on geometric mean concentrations from spot sample distributions is likely to underestimate the geometric mean of daily or longer-term average concentrations and therefore the geometric mean of dose rates for very rapidly eliminated analytes. Finally, the use of creatinine-adjusted concentrations provided better concordance among various metrics than unadjusted concentrations.

Within-Individual Variation in Analyte Concentrations

Spot sample concentrations within individuals ranged over approximately two orders of magnitude for most analytes examined here, except for triclosan, in which the within-individual variation covered just over one order of magnitude. The 24 h composite concentrations within individuals tend to vary by less than a factor of 10 over the course of multiple days, except for MEHHP (Figures 3 and 4). The day-to-day variation was smallest for triclosan (variation generally within a factor of 3) and largest for MEHHP (variation of more than 10-fold over the course of several days). The 24 h composite concentrations can be regarded as reflective of daily exposure amounts. This gives a picture of within-individual variation in daily exposure rates to these compounds in the individuals studied here.

Intraclass Correlation Coefficients

The ICCs calculated in this data set can be examined in light of information regarding exposure pathways and pharmacokinetics for the target analytes. Not surprisingly, the ICCs are lower for those chemicals with food as the predominant exposure pathway — BPA and MEHHP, the DEHP metabolite — compared with those for which exposure is associated with the use of specific products (triclosan, MEP, methyl paraben). This reflects larger between-individual variation that results from differences in product use as well as more episodic use patterns for products compared with food ingestion. In addition, half-life of elimination has a role in these patterns as well. For compounds with half-lives that are relatively short compared with the interval between exposure episodes, greater swings in urinary concentration between exposure events would be predicted within an individual, with a greater proportion of the concentration *vs* time curve at levels below the arithmetic mean owing to the shape of the first-order elimination curve. Given that the ICC is calculated as the ratio of between-person variability and the sum of between-person and within-person variability, and given that within-person variability is inversely dependent on half-life (the shorter the half-life, the greater the swings in urinary concentrations within a day), all things being equal, the ICC will be positively correlated with half-life.

These ICCs can be compared with ICCs for the same analytes from previous studies. Some of these studies are summarized in Aylward et al.[1] (see Table 5 in that publication). Previous studies have found mixed results for MEP, with both relatively low[21,22] and relatively high[23–25] ICCs reported. Numerous studies have reported relatively low ICCs for BPA (summarized in ref. 1). Teitelbaum et al.[22] reported moderate ICCs for triclosan. None of these studies conducted intensive sampling over a period of a few days similar to that presented for the CDC and CEFIC data sets, so some of the differences in ICCs may relate to longer-term differences in exposure levels and other factors reflected in the different populations and sampling regimens represented by these studies.

Approaches Based on Consideration of Pharmacokinetics and Extrapolation from Other Studies

Several groups have attempted to assess the degree of correspondence between distributions of spot sample

concentrations collected from a diverse population and the distribution of some longer-term average concentration.[1,26] Aylward et al.[1] assessed the impact of temporal variability and the impact in overestimation in urine concentrations in spot samples as compared with 24 h averages. The controlling factors were half-life of elimination and frequency of exposure. Knowing these two factors, some estimates could be made in how spot and 24 h distributions might differ.

Pleil and Sobus[26] have developed a mathematical approach to using ICCs to convert a distribution of spot samples into a likely distribution of longer-term averages. Their approach relies on an assumption that the spot sample distribution conforms to a lognormal distribution, that longer-term average concentrations can be represented by a lognormal distribution, and that ICCs from one studied population are applicable to other populations. If these conditions are generally met, the method may allow a more generalized set of predictions based on spot samples.

Factors Influencing Variation

As noted above, the ICCs provide indications of the inter- *vs* intra-individual variability and support what would be expected on the basis of the likely source and frequency of exposure for a compound. The analytes for which data were available for these analyses provide a relatively narrow range of half-lives. However, as half-life increases relative to typical exposure intervals, the degree of variation in spot sample concentrations will tend to more closely resemble the degree of variation in longer-term average concentrations in an individual.[1]

For the compounds for which exposures occur primarily as a result of widespread food contamination (e.g., BPA, MEHHP/DEHP),[27] the variation between individuals is relatively small and the variation within an individual is generally greater than the variation between individuals (low ICC — Table 3). Conversely, for the compounds for which exposure results from the use of certain consumer products such as facial creams, certain brands of soap, toothpaste, and mouthwash (e.g., triclosan, MEP), the variation between individuals can be quite large and the variability within an individual is much less than the overall population variation (high ICC — Table 3).[12]

Another issue potentially contributing to population variation in biomarker concentrations relates to the variability in urinary excretion fraction estimated for some compounds. In particular, in a controlled-dosing study for triclosan, the measured urinary excretion fraction varied widely, from 20% to 80%.[20] Such variation in excretion fraction may contribute to the overall population variation in biomarker concentrations observed and contributes to uncertainty in reconstructed exposure rates using reverse dosimetry approaches.

Other Sources of Systematic Variation

The data sets analyzed here do not allow empirical evaluation of several other issues of interest. Because these two data sets are focused on a relatively short window of exposure time, they do not provide information for individuals on how representative the distributions of spot, 24 h composite, or multiday average concentrations are for longer-term average concentrations. That is, if an individual's exposure patterns differ over weeks or between seasons, that variation is not necessarily represented by the variations presented here. However, the relationship between the degree of variation between spot and 24 h composite samples may be similar if the routes of exposure and relative frequencies of exposure are similar, even if the absolute levels are different.

These data sets also do not allow an assessment of whether variation is a function of age or conditions like pregnancy or BMI, because the individuals sampled intensively in these two studies represent a small number of individuals with a limited range of characteristics. Systematic variation in physiological factors such as urinary flow rates,[28] creatinine excretion,[8,29] and other factors may also influence population variation in sampled concentrations.[30] Awareness of these issues can improve the use and interpretation of existing biomonitoring data and the design of future biomonitoring studies.

Value of Intensive Serial Urine Collection Studies

The extensive data available in the CDC and CEFIC studies provide a robust basis to allow calculation of ICCs and compare percentiles between the spot, 24 h, and longer-term average distributions for these two populations over the sampled time periods. Extension of the results observed here to interpretation of population-based biomonitoring studies in which a single spot sample was collected from many more participants requires some caution, however, the data sets do provide insight into the properties of biomarkers for rapidly eliminated compounds within individuals and the potential implications for interpretation of the larger population studies. There are numerous studies reporting ICCs for chemicals in which the study population provided only a few urine voids over either short or extended periods of time (see Aylward et al.[1] for a review of available studies). However, the ICCs calculated from only a limited number of urine voids (or blood samples) may be of limited value for making the types of assessments included in this report. Additional studies like that of the CDC and CEFIC studies will aid in assessments for additional compounds and allow broader conclusions based on half-life and types/sources of exposures.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The analyses presented here depended upon the availability of the high quality, intensive data sets from the CDC and from the CEFIC Long-Range Research Institute project "CEFIC-HBM4: Understanding inter- and intra-individual variability in HBM spot samples," and we thank CDC and CEFIC-LRI for the use of these data sets. Data analysis and manuscript preparation was funded by Health Canada under Contract 4500324122.

REFERENCES

1 Aylward LL, Kirman CR, Adgate JL, McKenzie LM, Hays SM. Interpreting variability in population biomonitoring data: role of elimination kinetics. *J Expo Sci Environ Epidemiol* 2012; **22**: 398–408.
2 Preau JL Jr, Wong LY, Silva MJ, Needham LL, Calafat AM. Variability over 1 week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among eight adults: an observational study. *Environ Health Perspect* 2010; **118**: 1748–1754.
3 Calafat AM, McKee RH. Integrating biomonitoring exposure data into the risk assessment process: phthalates [diethyl phthalate and di(2-ethylhexyl) phthalate] as a case study. *Environ Health Perspect* 2006; **114**: 1783–1789.
4 Government of Canada. Proposed Approach for Cumulative Risk Assessment of Certain Phthalates Under the Chemicals Management Plan. *Health Canada, Environment Canada* 2015. http://www.chemicalsubstanceschimiques.gc.ca/group/phthalate/index-eng.php-a1.
5 Government of Canada. Preliminary Assessment - Triclosan. *Health Canada, Environment Canada* 2012. http://www.ec.gc.ca/ese-ees/default.asp?lang=En&n=6EF68BEC-1.
6 Lakind JS, Naiman DQ. Daily intake of bisphenol A and potential sources of exposure: 2005-2006 National Health and Nutrition Examination Survey. *J Expo Sci Environ Epidemiol* 2011; **21**: 272–279.
7 Mage DT, Allen RH, Gondy G, Smith W, Barr DB, Needham LL. Estimating pesticide dose from urinary pesticide concentration data by creatinine correction in the Third National Health and Nutrition Examination Survey (NHANES-III). *J Expo Anal Environ Epidemiol* 2004; **14**: 457–465.
8 Mage DT, Allen RH, Kodali A. Creatinine corrections for estimating children's and adult's pesticide intake doses in equilibrium with urinary pesticide and creatinine concentrations. *J Expo Sci Environ Epidemiol* 2008; **18**: 360–368.

9 United States Environmental Protection Agency. (USEPA). Reregistration Eligibility Decision for Triclosan. Office of Prevention, Pollution, and Toxic Substances. EPA 739-RO-8009. Available at http://www.epa.gov/oppsrrd1/reregistration/REDs/2340red.pdf. 2008.
10 Li Z, Romanoff LC, Lewin MD, Porter EN, Trinidad DA, Needham LL *et al.* Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. *J Expo Sci Environ Epidemiol* 2010; **20**: 526–535.
11 Ye X, Wong LY, Bishop AM, Calafat AM. Variability of urinary concentrations of bisphenol A in spot samples, first morning voids, and 24-hour collections. *Environ Health Perspect* 2011; **119**: 983–988.
12 Koch HM, Aylward LL, Hays SM, Smolders R, Moos RK, Cocker J *et al.* Inter- and intra-individual variation in urinary biomarker concentrations over a 6-day sampling period. Part 2: personal care product ingredients. *Toxicol Lett* 2014; **231**: 261–269.
13 Smolders R, Koch HM, Moos RK, Cocker J, Jones K, Warren N *et al.* Inter- and intra-individual variation in urinary biomarker concentrations over a 6-day sampling period. Part 1: metals. *Toxicol Lett* 2014; **231**: 249–260.
14 van Haarst EP, Heldeweg EA, Newling DW, Schlatmann TJ. The 24-h frequency-volume chart in adults reporting no voiding complaints: defining reference values and analysing variables. *BJU Int* 2004; **93**: 1257–1261.
15 Volkel W, Kiranoglu M, Fromme H. Determination of free and total bisphenol A in human urine to assess daily uptake as a basis for a valid risk assessment. *Toxicol Lett* 2008; **179**: 155–162.
16 Moos RK, Angerer J, Bruning T, Koch HM. Human metabolism and excretion kinetics of methyl, n- and iso- butyl parabens after oral dosage. *Poster presented at the International Society of Exposure Science Annual Meeting October 2014* at Cincinatti, Ohio, USA.
17 Aylward LL, Hays SM, Gagne M, Krishnan K. Derivation of biomonitoring equivalents for di(2-ethylhexyl)phthalate (CAS No. 117-81-7). *Regul Toxicol Pharmacol* 2009; **55**: 249–258.
18 Koch HM, Bolt HM, Preuss R, Angerer J. New metabolites of di(2-ethylhexyl) phthalate (DEHP) in human urine and serum after single oral doses of deuterium-labelled DEHP. *Arch Toxicol* 2005; **79**: 367–376.
19 Krishnan K, Gagne M, Nong A, Aylward LL, Hays SM. Biomonitoring equivalents for triclosan. *Regul Toxicol Pharmacol* 2010; **58**: 10–17.
20 Sandborgh-Englund G, Adolfsson-Erici M, Odham G, Ekstrand J. Pharmacokinetics of triclosan following oral ingestion in humans. *J Toxicol Environ Health A* 2006; **69**: 1861–1873.
21 Adibi JJ, Whyatt RM, Williams PL, Calafat AM, Camann D, Herrick R *et al.* Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. *Environ Health Perspect* 2008; **116**: 467–473.
22 Teitelbaum SL, Britton JA, Calafat AM, Ye X, Silva MJ, Reidy JA *et al.* Temporal variability in urinary concentrations of phthalate metabolites, phytoestrogens and phenols among minority children in the United States. *Environ Res* 2008; **106**: 257–269.
23 Hauser R, Meeker JD, Park S, Silva MJ, Calafat AM. Temporal variability of urinary phthalate metabolite levels in men of reproductive age. *Environ Health Perspect* 2004; **112**: 1734–1740.
24 Hoppin JA, Brock JW, Davis BJ, Baird DD. Reproducibility of urinary phthalate metabolites in first morning urine samples. *Environ Health Perspect* 2002; **110**: 515–518.
25 Peck JD, Sweeney AM, Symanski E, Gardiner J, Silva MJ, Calafat AM *et al.* Intra- and inter-individual variability of urinary phthalate metabolite concentrations in Hmong women of reproductive age. *J Expo Sci Environ Epidemiol* 2010; **20**: 90–100.
26 Pleil JD, Sobus JR. Estimating lifetime risk from spot biomarker data and intraclass correlation coefficients (ICC). *J Toxicol Environ Health A* 2013; **76**: 747–766.
27 Aylward LL, Lorber M, Hays SM. Urinary DEHP metabolites and fasting time in NHANES. *J Expo Sci Environ Epidemiol* 2011; **21**: 615–624.
28 Hays SM, Aylward LL, Blount BC. Variation in urinary flow rates according to demographic characteristics and body mass index in NHANES: potential confounding of associations between health outcomes and urinary biomarker concentrations. *Environ Health Perspect* 2015; **123**: 293–300.
29 Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. *Environ Health Perspect* 2005; **113**: 192–200.
30 Aylward LL, Hays SM, Smolders R, Koch HM, Cocker J, Jones K *et al.* Sources of variability in biomarker concentrations. *J Toxicol Environ Health B Crit Rev* 2014; **17**: 45–61.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

© The Author(s) 2017