



Published in final edited form as:

Environ Int. 2019 January ; 122: 222–230. doi:10.1016/j.envint.2018.11.012.

Variability of Urinary Concentrations of Phthalate Metabolites during Pregnancy in First Morning Voids and Pooled Samples

Hyeong-Moo Shin^{a,b,*}, Deborah H. Bennett^a, Jacqueline Barkoski^a, Xiaoyun Ye^c, Antonia M. Calafat^c, Daniel Tancredi^d, and Irva Hertz-Picciotto^{a,e}

^aDepartment of Public Health Sciences, University of California, Davis, California, USA

^bDepartment of Earth and Environmental Sciences, University of Texas, Arlington, Texas, USA

^cCenters for Disease Control and Prevention, Atlanta, Georgia, USA

^dDepartment of Pediatrics, University of California, Davis, California, USA

^eUC Davis MIND (Medical Investigations of Neurodevelopmental Disorders) Institute, Sacramento, California, USA

Abstract

Background: Because phthalates are quickly metabolized and excreted in urine, and human exposures tend to be episodic, phthalate metabolite concentrations measured in a maternal spot urine sample are only indicative of recent exposure.

Objective: To examine temporal variability of pregnant women's phthalate exposure using multiple first morning voids (FMV) and pooled samples.

Methods: We quantified 14 metabolites of eight phthalates in 577 urine samples collected from 188 pregnancies in the MARBLES (Markers of Autism Risk in Babies – Learning Early Signs) study. We calculated intraclass correlation coefficients (ICCs) using two samples of the same urine type (i.e., two FMVs or two pools) collected across the 2nd and 3rd trimesters. We also calculated ICCs and FMV/pool concentration ratios using two samples (i.e., two FMVs or one FMV and one pool) collected within the same trimester.

Results: Overall, ICCs were higher in pooled samples (0.24 – 0.87) than in FMVs (0.08 – 0.69). Regardless of the sample type, ICCs tended to be higher for metabolites for which exposure sources are personal care products or indoor residential materials than those for which diet is an important exposure source. ICCs tended to increase and FMV/pool ratios tended to decrease with an increasing number of composite samples in the pools.

*Corresponding author (present address): Hyeong-Moo Shin, Ph.D. University of Texas, Arlington, 500 Yates Street, Box 19049, Arlington, Texas 76019, hyeongmoo.shin@uta.edu Voice: 817-272-2970, Fax: 817-272-2628.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclaimer: The authors declare they have no actual or potential competing financial interests.

Conclusions: Our study helped determine the number of samples needed to capture moderate to high reproducibility of individual's average exposure to phthalates and the average exposure can be differently characterized depending on the number of samples in the pools.

Keywords

first morning void; intraclass coefficient; phthalate exposure; pooled sample; variability

INTRODUCTION

Phthalates are ubiquitous chemicals used in personal care products such as cosmetics, fragrances, and shampoos, and can be present in indoor residential materials such as polyvinyl chloride flooring, plastics, children's toys, vinyl tiles, and shower curtains (Dodson et al. 2012; Hauser and Calafat 2005; Heudorf et al. 2007). Human exposure to phthalates can occur via various exposure pathways and routes (Schettler 2006). Personal care products or indoor residential materials are known to be principal exposure sources for low molecular weight phthalates (e.g., diethyl phthalate, DEP) and diet is known to be a primary exposure source for high molecular weight phthalates (e.g., di(2-ethylhexyl) phthalate, DEHP) (Koch et al. 2013; Rudel et al. 2011; Varshavsky et al. 2018; Wormuth et al. 2006).

Urinary phthalate metabolites have been used extensively as biomarkers of phthalate exposure (Koch and Calafat 2009). Phthalates are quickly metabolized and excreted in urine, with elimination half-lives on the order of hours (Hoppin et al. 2002). Thus, phthalate metabolite concentrations measured in a spot urine sample reflect recent exposure (Fromme et al. 2007; Hauser et al. 2004; Preau et al. 2010). Nevertheless, phthalate metabolites have been detected in most of the urine samples from a representative sample of the U.S. general population (CDC 2017), suggesting widespread exposures to phthalates. Phthalate biomarkers have also been detected in the urine of pregnant women (Adibi et al. 2008; Cantonwine et al. 2014; Enke et al. 2013; Ferguson et al. 2014a; Huang et al. 2009; Lin et al. 2011; Meeker et al. 2009; Woodruff et al. 2011), suggesting the possibility of gestational exposures.

Of interest, several studies have found weak to moderate reproducibility of phthalate metabolite concentrations throughout pregnancy (Adibi et al. 2008; Braun et al. 2012; Cantonwine et al. 2014; Ferguson et al. 2014a; Fisher et al. 2015), which may lead to exposure misclassification when characterizing average exposures based on a single or a few samples. Collecting multiple urine specimens from the same subject could capture a person's variability in urinary concentrations and thus better characterize average exposures during pregnancy (Perrier et al. 2016). However, increasing the number of urine samples per subject could increase the analytical cost and limit the number of subjects to be included for epidemiologic analyses. Alternatively, within-subject pools comprised of multiple samples have been suggested for use in epidemiologic studies of chemicals with short elimination half-lives (Calafat 2016; Heffernan et al. 2014; Perrier et al. 2016). Pooling multiple biological specimens within each subject can both increase the number of specimens analyzed per subject and decrease analytical cost as when pooling specimens between

subjects within a given population (Mitchell et al. 2014; Vexler et al. 2006). Moreover, use of pooled samples could decrease bias from measurement error and increase statistical power of epidemiologic studies that have collected multiple samples longitudinally (Perrier et al. 2016).

For the present study, we characterized variability of phthalate exposure using a total of 577 urine samples collected from 188 pregnancies. Specifically, we quantified phthalate metabolite concentrations in urine samples (1) to evaluate temporal variability of urinary concentrations of phthalate metabolites within the same trimester and across two trimesters and (2) to evaluate whether within-subject pools can improve phthalate exposure characterization. To our knowledge, this is the first study to assess the usefulness of multiple first morning voids (FMVs) and within-subject pools to characterize variability of phthalate exposures during pregnancy. We expect that a comprehensive understanding of variability of phthalate exposure from the present study will improve the characterization of pregnant women's exposure to phthalates.

METHODS

Study population

This present study was nested within the MARBLES (Markers of Autism Risk in Babies – Learning Early Signs) study. MARBLES is a cohort study following pregnant women who previously delivered a child who developed autism spectrum disorder (ASD) and therefore are at high risk for delivering another infant who will also develop ASD (Ozonoff et al. 2011). MARBLES families are primarily recruited from lists of children receiving services for autism through the California Department of Developmental Services, as well as from other studies, by self- or provider referrals, and obstetrics/gynecology clinics. Details of study design, recruitment, eligibility, sample size, exposure data, and developmental diagnosis are available elsewhere (Hertz-Picciotto et al. in press).

For the present study, we selected 178 mothers who provided FMVs and/or 24-hour urine samples during pregnancy collected from 01/2007 to 02/2014. Among 178 mothers, 8 mothers participated in the study for two different pregnancies and one mother participated for three different pregnancies over four years. All urine samples included in the present study were collected from a total of 188 unique pregnancies. Basic demographic information including age, race, and education of the selected mothers and the entire population of mothers enrolled in MARBLES is described in the Supplemental Material (see Table S1).

This study was approved by the institutional review boards for the State of California and the University of California Davis (UC Davis). The analysis of blinded specimens by the Centers for Disease Control and Prevention (CDC) laboratory was determined not to constitute engagement in human subjects' research. Participants provided written informed consent before collection of any data.

Urine sample collection

Each woman in the MARBLES study was instructed to collect three FMVs (taken one week apart) and one 24-hour urine sample in each trimester, placing each sample collected prior to

the day of her visit in their home freezer, allowing us to examine temporal variability of biomarkers of environmental chemicals with short elimination half-lives. Women freely picked a day of the week and were asked to do the collection on that day each week for 3 consecutive weeks (29.3% were collected during the weekend and 70.7% were collected on weekdays). One week after the last FMV collection, women were instructed to collect the 24-hour urine sample on the same day of the week as the FMV samples. On the day the collection of the 24-hour urine sample was to start, women were asked to empty their bladder at 8:00 AM, discard this urine, and collect all urine they passed over a 24-hour period, including the first void the following morning. Women were instructed to collect all urine in a clean plastic container (e.g., measuring jug or urine hat), pour in the labeled bottle provided by the study staff, and keep the collection bottle in the refrigerator or other cool place. Samples were returned to the laboratory at UC Davis, thawed, aliquoted, and then stored at -80°C at the UC Davis biorepository.

Sample preparation for chemical analysis

Although almost all of the mothers included in the current study ($n = 178$) provided urine samples during the 2nd and 3rd trimesters, only 72 mothers provided urine samples during the 1st trimester and most of them provided a single urine sample during the 1st trimester. Thus, urine samples collected during the 2nd and 3rd trimesters were only used in the current study. Although all mothers were asked to collect four samples (three FMVs and one 24-hour sample) per trimester, many did not. We further selected nine mothers who provided at least four (either FMV or 24-hour) samples for each of the 2nd and the 3rd trimesters, and analyzed their samples individually to study variability. To reduce analytical cost, for mothers who provided three or more urine specimens within a trimester, we selected the first FMV as an individual sample to be comparable with studies that collected only 1 urine sample and pooled all remaining samples (including 24-hour samples if any) for that trimester to best capture average exposure. To pool urine specimens, 5 ml of each specimen was transferred by sterile pipette into a sterile 20 ml conical tube (or 50 ml conical tube if more than 4 specimens were pooled) and combined using a vortex mixer for 10 seconds. Next, 2 ml aliquots of the pooled specimen were transferred by pipette into sterile vials and then shipped to the CDC laboratory (Barkoski et al. In press). The remaining aliquots were stored in -80°C freezers at the UC Davis biorepository.

After pooling at the UC Davis laboratory, a total of 577 urine samples remained for analysis: 362 FMVs, 17 24-hour samples (from nine mothers for the graphical representation of the concentration variability over the course of the 2nd and 3rd trimesters) and 198 within-subject pools (~78% of the pools were comprised of at least one 24-hour sample). The number of samples included in each of the 198 pools varied based on the number of samples each woman collected. The type and number of urine samples collected and analyzed in this study are summarized in Figure 1.

Urinary phthalate metabolite quantification

We shipped the urine samples in 2 mL aliquots to the CDC for analysis. At CDC, we quantified the urinary concentrations for 14 metabolites of eight phthalates (see Table S2 for parent compounds of each metabolite) using online solid phase extraction coupled with

high-performance liquid chromatography with isotope dilution-tandem mass spectrometry as described elsewhere (Silva et al. 2007). The analytical measurements followed strict quality control/quality assurance protocols, including participation in quality assessment schemes to demonstrate the method accuracy and precision, as required by the Clinical Laboratory Improvement Act of 1988 (CLIA '88) certification. Furthermore, along with study samples, each analytical run included spiked quality control (QC) materials and reagent blanks. For this study, depending on the analyte and QC concentration, the average relative percent difference (RPD) of 59 repeated measures of these QCs was 2.6–8.5%. In addition, blind replicate analyses of 51 individual pairs of duplicate study samples showed good agreement: average RPD was 7% (range: 4–15%, depending on the analyte). These results assure the accuracy and reliability of the presented data. Details of the analytical procedure used are available at https://www.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/PHTHTE_H_MET_Phthalates.pdf.

The fourteen phthalate metabolites quantified were: monoethyl phthalate (MEP), mono-isobutyl phthalate (MiBP), monohydroxy-isobutyl phthalate (MHiBP), mono-n-butyl phthalate (MBP), monohydroxy-n-butyl phthalate (MHBP), monobenzyl phthalate (MBzP), mono(2-ethylhexyl) phthalate (MEHP), mono(2-ethyl-5-hydroxyhexyl) phthalate (MEHHP), mono(2-ethyl-5-oxohexyl) phthalate (MEOHP), mono(2-ethyl-5-carboxypentyl) phthalate (MECPP), mono(3-carboxypropyl) phthalate (MCP), mono-isononyl phthalate (MNP), mono-carboxyisooctyl phthalate (MCOP), and mono-carboxyisononyl phthalate (MCNP). For metabolite concentrations below the limit of detection (LOD, Table 1), we assigned a value of LOD divided by the square root of 2 (Hornung and Reed 1990).

Correction for urinary dilution

To account for urinary dilution (Adibi et al. 2008; Hauser et al. 2004), we measured specific gravity (SG) in each sample (either individual or pooled) using a digital handheld refractometer (Atago Co., Ltd., Tokyo, Japan) at UC Davis. SG for all samples included in this study ranged from 1.001 to 1.033. Summary statistics for SG of all samples, FMVs and pools are provided in the Supplemental Material (see Table S3). Urinary phthalate metabolite concentrations were corrected for dilution using the following formula: $C_{SG} = C [(1.012 - 1)/(SG - 1)]$, where C_{SG} is the SG-corrected metabolite concentration (in micrograms per liter ($\mu\text{g/L}$)), C is the measured concentration in urine (in $\mu\text{g/L}$), 1.012 is the median SG of all samples analyzed, and SG is the specific gravity of each sample (Cantonwine et al. 2014; Ferguson et al. 2014b; Meeker et al. 2009). With an SG-correction factor, the computed value in the bracket of the above formula, outside the range of 0.5 and 2.0, the SG-corrected concentrations might not be reliable (Rosenberg et al. 1989). Thus, we trimmed correction factors below this range to 0.5 and correction factors above this range to 2.0. More than 85% of the samples had an SG-correction factor between 0.5 and 2.0. We used SG-corrected metabolite concentrations for all statistical analyses.

Statistical analysis

To assess within-subject variability, we calculated the intraclass correlation coefficient, ICC, defined as the ratio of between-subject variance to total variance (= within-subject variance + between-subject variance). ICC ranges from 0 (no reproducibility) to 1 (perfect

reproducibility); ICC=1 means 100% of variance is due to between-subject differences (Adibi et al. 2008; Rosner 2000). To evaluate the effect of a sample type (FMVs or pools) on variability of phthalate metabolite concentrations, we calculated ICCs using two samples of the same urine type collected across the 2nd and 3rd trimesters: (1) two first FMVs ($n = 186$ from 93 mothers) and (2) two pooled samples ($n = 118$ from 59 mothers). Two samples of the same urine type (i.e., FMVs or pools) used in these ICC calculations were collected approximately 3 months apart. We did not calculate ICCs for 24-hour samples due to a small sample size ($n = 17$).

Each pooled sample had a varying number of composite samples, ranging from 2 to 4. To examine how much pooled samples with an increasing number of composites can improve reproducibility of individual's exposure, we also calculated ICCs using two samples collected within the same trimester for the following four combinations: (1) two FMVs ($n = 98$), (2) one FMV and one pooled sample with 2 composites ($n = 128$), (3) one FMV and one pooled sample with 3 composites ($n = 206$), and (4) one FMV and one pooled sample with 4 composites ($n = 44$). Two FMVs used in ICC calculations were collected approximately 1 week apart. The last sample in the pool and the first FMV were approximately 2, 3, or 4 weeks apart depending on the number of composites in the pool. The number of individual and pooled samples used to compute ICCs varied by mother, because the number of samples each woman collected for each trimester varied. Using the same dataset above, we also calculated the concentration ratio of an FMV to a pooled sample ($C_{\text{FMV}}/C_{\text{pool}}$). Lastly, we computed ICCs separately for each of the 2nd and 3rd trimesters to investigate trimester-specific temporal variability of phthalate metabolite concentrations.

Statistical analyses were performed using R version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) and MATLAB 2018a (The MathWorks, Inc., Natick, MA, USA). ICCs and 95% confidence intervals (CI) were estimated using variance component estimates from a one-way analysis of variance (ANOVA). We used the 'ICCest' function in R for these calculations. Ln-transformed SG-corrected metabolite concentrations were used in all ICC calculations. For mothers who participated in the study for multiple pregnancies ($n = 9$), we treated samples from each pregnancy independently.

RESULTS

Descriptive analyses

The distribution of SG-corrected phthalate metabolite concentrations from FMVs and pools collected from 188 unique pregnancies is summarized in Table 1. Except for MHBP (82%), MEHP (82%), and MNP (49%), the other 11 metabolites were detected in more than 90% of the samples; 10 of 14 metabolites were detected in more than 97% of the samples. Although this is not a direct comparison between two populations of pregnant mothers, we also compared geometric mean (GM) urinary concentrations of phthalate metabolites in the present study (pregnant mothers) to those in the U.S. general female population (pregnant and non-pregnant) reported in the National Health and Nutrition Examination Survey (NHANES) (CDC 2017). In general, the metabolite concentrations were comparable, but GM concentrations were slightly higher in this study than those reported in NHANES

(2011–2012). When restricting our comparison to the same study period (NHANES 2011–2012) and the concentrations in FMVs, GM concentrations of MBP and MBzP were approximately 40% higher in the present study than in NHANES. In the case of MEP, GM concentrations were approximately 40% lower in our study than in NHANES.

Variability within the same trimester

Overall, ICCs (Figure 2, Table S4) from the samples collected within the same trimester (2nd or 3rd) tended to be larger for mothers who provided three or more samples than for those who provided only two samples. For all metabolites, median ICCs for one FMV and pooled samples with 4 composites (the last pair) were significantly larger than those for two FMVs (the first pair) at the 5% level. In addition, metabolites with relatively consistent exposure sources such as personal care products or indoor residential environments (e.g., MEP, MiBP, MHiBP, MBP, MHBP, MBzP) tended to have higher ICCs and narrower confidence intervals (assuming the same number of composites in each pooled sample), compared to metabolites with relatively variable exposure sources mainly through diet (e.g., MEHP, MEHHP, MEOHP, MECPP, MCPP, MNP, MCOP, MCNP) (Varshavsky et al. In press). Compared to ICCs from two FMVs, ICCs for the former group of metabolites increased (e.g., 15% for MHBP up to 100% for MiBP) when the pooled sample was comprised of 3 composites, except for MEP (Figure 2A). On the other hand, ICCs for the latter group of metabolites increased (e.g., 33% for MEHP up to 500% for MCNP) when the pooled sample was comprised of 4 composites (Figure 2B). For the latter group of metabolites and MEP, we also observed that ICCs computed with one FMV and one pooled sample with 3 composites (the third pair) were smaller than those with one FMV and one pooled sample with 2 composites (the second pair).

ICCs were separately computed using samples collected for each of the 2nd and 3rd trimesters (Figure S1 and Figure S2). For some metabolites and pairs, median ICCs were significantly different between the 2nd and 3rd trimesters at the 5% level and confidence intervals became large for all metabolites due to the reduced number of sample size. Trimester-specific ICCs did not follow the same trend as those when the sample pairs from both trimesters were combined (Figure 2). However, for all metabolites, median ICCs for one FMV and pooled samples with 3 or 4 composites were consistently larger than those for two FMVs, regardless of the trimester.

From the same dataset as above, the concentration ratio of an FMV to a pooled sample indicated that median ratios decreased with an increasing number of samples, especially for metabolites having a large portion of exposure sources from diet (Figure 3B, Table S5). For metabolites having no significant exposure source from diet (Figure 3A), median ratios tended to decrease, but were relatively close to 1.

Variability across the 2nd and 3rd trimesters

From the samples collected across the 2nd and 3rd trimesters (Figure 4), ICCs were higher in pooled samples (0.24 – 0.87) than in FMVs (0.08 – 0.69). Except for MEP, all metabolites were statistically significant at the 5% level (i.e., p-values for a two sample *t*-test were less than 0.05). Regardless of the sample type, ICCs tended to be higher for metabolites for

which exposure sources are mostly personal care products or indoor surface materials (Figure 4A) than metabolites for which diet is a primary exposure source (Figure 4B). Specifically, in the pooled samples, ICCs for the former group of metabolites ranged from 0.55 to 0.87, indicating high reproducibility (i.e., lower temporal variability). On the other hand, ICCs for the latter group of metabolites ranged from 0.24 to 0.56, indicating low reproducibility (i.e., greater temporal variability). DEHP metabolites showed relatively moderate reproducibility (ICC = 0.51 – 0.56).

From the graphical representation of the urinary concentrations of the sum of DEHP metabolites (DEHP) and of MEP over the 2nd and 3rd trimesters for 9 mothers (Figure 5), we found significant differences in metabolite concentrations across sample collection times for some of the subjects. For DEHP, four mothers (Figure 5A) showed relatively low within-subject variability with coefficient of variation (CV) less than 0.45. Thus, DEHP was almost consistently higher in the order of subject 2, 1, 3, and 4. On the other hand, five mothers (Figure 5B) showed high within-subject variability with CV greater than 0.95. The urinary concentrations for subject 9 varied almost two orders of magnitude during the sampling period. Three of four mothers who had low within-subject variability in DEHP (subjects 1, 3, 4) showed high within-subject variability for MEP, with CV ranging from 0.71 to 1.69 (Figure 5C). Three of five mothers who had high within-subject variability in DEHP (subjects 5, 6, 8) showed low within-subject variability for MEP, with CV less than 0.67 (Figure 5D). These results suggest that the degree of within-subject variability varies not only between participants, but also by metabolites within participants.

DISCUSSION

Results from this study support that within-subject pooling can lead to more reproducible individual's exposure (as measured by ICCs) than FMVs. Overall, we observed higher ICCs from pooled samples than from FMVs, except for MEP. Specifically, for metabolites with relatively variable exposure sources mainly through diet (Figure 4B), ICCs of pooled samples were almost two times larger than those of FMVs. Our finding is consistent with a recent study showing that the average urinary flame retardant metabolite concentrations across pregnancy had higher agreement (measured by ICC) with the pooled urine samples than with individual samples (Romano et al. 2017). We also observed higher ICCs (Figure 2A) for pooled samples with 3 or 4 composites, compared to two FMVs and one FMV with one pool with 2 composites. For metabolites with relatively variable exposure sources mainly through diet (Figure 2B), we also observed that ICCs computed with one FMV and one pooled sample with 3 composites were smaller than those with one FMV and one pooled sample with 2 composites. This is partly due to the relatively high percentage of 24-hour samples in the pool with 3 composites (90%), compared to that in the pool with 2 composites (59%). For this group of metabolites, urinary concentrations in FMVs after overnight fasting and those in 24-hour samples may be much different. In addition, we observed that, in general, metabolites with relatively consistent exposure sources such as personal care products or indoor residential environments had relatively high ICCs, compared to metabolites with relatively variable exposure sources mainly through diet. This pattern was also observed in previous studies and may reflect the known differences in exposure sources among phthalates (Preau et al. 2010; Zota et al. 2014).

A large number of FMVs and pooled samples allowed us to evaluate temporal variability of urinary concentrations of phthalate metabolites. We used a total of 927 urine samples (752 FMV and 175 24-hr samples) collected from a subset of MARBLES pregnant mothers and after pooling, analyzed 577 samples. A total of 75 individual samples from nine women were analyzed to evaluate graphically the concentration variability over the course of the 2nd and 3rd trimesters. From these longitudinal samples, we observed that five mothers had urinary concentrations that varied dramatically (as much as almost two orders of magnitude) during the 2nd and 3rd trimesters depending on the subject and the metabolite. For such women, using one or a few spot or FMV or 24-hour samples collected at different time points may lead to bias in the estimated dose-response relationship toward the null, and may reduce statistical power to detect such an association (Perrier et al. 2016).

Approximately 300 pairs of two FMVs or one FMV and one pooled sample with a varying number of composites collected within the same trimester allowed us to determine the number of samples required to characterize pregnant women's average phthalate exposures. Because of this rich dataset, we also observed that exposure reproducibility (measured by ICCs) for women who provided more samples (more compliant) differed from those who provided less samples (less compliant). We note that physically combining equal volume aliquots from multiple samples leads to an arithmetic average of the constituents. On the other hand, most environmental epidemiologic studies work on the log scale when creating a summary measure of multiple individually measured biospecimens to avoid pulling the summary measure towards outliers in the right tail of the distribution. Because of the impact of outliers, the average concentration in the pool tended to be larger with an increasing number of samples than the concentration of a single FMV, especially for phthalate metabolites for which diet is an important exposure source (Figure 3B, Table S5). This indicates that for two mothers who had similar average daily exposures on the day of urine collection, an individual's long-term average exposure characterized from a given number of samples will vary depending on the number of samples each provided, and the concentration from pooled samples diluting an outlier will be a more representative individual's exposure than that from a single FMV with an exceptionally high concentration, an outlier. Thus, this finding may provide appropriate guidance for epidemiologists who are interested in within-subject pooling for biomarkers with a lognormal distribution. For example, when averaging metabolite concentrations by subject using one individual and one pooled sample with a varying number of composite samples, different weight needs to be accounted for between individual and pooled samples based on the number of samples per subject and the degree of reproducibility (e.g., ICC). Specifically, the within-subject pool with multiple composites needs to be weighted more heavily than a single spot or FMV sample. Moreover, for phthalate metabolites with relatively small within-subject variability (i.e., high ICCs), the pooled sample needs to be weighted less than metabolites with large within-subject variability (i.e., low ICCs). For example, if one phthalate metabolite has an ICC of 0.9, a small number of samples would be required to characterize an individual's average exposure. For this case, the contribution of the pooled sample to the average exposure would not be as great as that of the pooled sample for metabolites with an ICC of 0.1, which require a larger number of samples to characterize an individual's average exposure.

To our knowledge, this is the first study to assess whether pooled samples can improve phthalate exposure classification compared to one individual sample. Because urinary concentrations of phthalate metabolites vary over time, multiple samples are required to characterize the actual average urinary concentration throughout the time period. For pooled samples, the greater the number of individual samples pooled, the closer the estimate will be to the average urinary concentration. For metabolites with relatively consistent exposure sources such as from indoor environments or personal care products (Figure 2A), we could capture a high degree of reproducibility (ICC of 0.6 or higher) from four samples per subject (one individual sample and a pooled sample with 3 composite samples). For biomarkers of other phthalates with variable exposure sources such as from diet (Figure 2B), we were not able to achieve the same degree of reproducibility with even up to 5 samples, though having more samples definitely increases credibility of the assessed exposure. Thus, further studies with more than 5 samples are needed for these compounds to determine the number of samples to capture moderate to high reproducibility of individual's exposure (e.g., ICC of 0.6 or higher).

The samples in the present study were largely FMVs (70% of the samples analyzed), collected each week for up to three weeks in both the 2nd and 3rd trimesters. Although the concentrations of phthalate metabolites in multiple FMVs may increase reproducibility of individual's exposure, the average urinary concentration characterized from the majority of multiple FMVs might not necessarily represent individual's daily average exposure because of over-night fasting. Multiple spot samples collected from any time of a day could better represent individual's average exposure, but might reduce reproducibility of individual's exposure (Ye et al. 2011). Multiple 24-hour samples might minimize exposure misclassification, compared to multiple FMVs or spot samples, but is generally not feasible for large epidemiologic studies. This means that the actual average exposure we want to characterize might be different depending on the sample type (i.e., FMV versus spot versus 24-hour) as well.

The ICCs calculated from the participants of this study may not be representative of other populations or other exposure windows/sampling designs. The use of largely FMVs, which were collected a long time after the last meal, is likely to underestimate average exposure for phthalates with dietary sources (e.g., DEHP). However, our study participants had higher DEHP metabolite concentrations than NHANES women (using spot samples collected during the daytime) for the same time period (Table 1). Together, these findings suggest that women in our study had different exposure patterns on average as compared with U.S. averages and thus may have greater between-subject variance in exposure, which boosts ICCs.

CONCLUSIONS

For chemicals with short elimination half-lives such as phthalates, a key step when using concentrations of biomarkers for exposure assessment is to have a good understanding of within- subject variability in biomarker concentrations to determine the optimum sampling and analysis strategy. Our results suggest that pooling multiple individual samples from each subject will likely decrease bias from measurement error in analyses of associations between

phthalate biomarker concentrations and health outcomes. We observed that phthalate biomarker measurements characterized from pools with multiple urine samples demonstrably improved credibility of estimated exposures over the use of one or a few samples. We also learned that urinary concentration temporal patterns vary considerably from person to person even for metabolites of phthalates with relatively consistent daily exposure sources, likely reflecting differences in the major sources of exposure to these compounds among our study participants. In addition, we learned that the average concentrations of specimens from women who were more compliant (collecting more samples) were systematically higher than those with fewer samples for most metabolites because of the impact of potential outliers included in a pooled sample. This finding could be important information for epidemiologists who are interested in within-subject pooling for biomarkers with a lognormal distribution or other skewed distributions. Environmental health scientists seeking to investigate health consequences from exposures to chemicals with short elimination half-lives would benefit from focusing attention on further strategies to minimize exposure measurement error.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank the MARBLES participants for helping make this research possible. This research was supported by grants from the National Institute of Environmental Health Sciences (R21-ES025551, R01-ES020392, R24-ES028533). We thank Manori Silva, Ella Samandar, Jim Preau, and Tao Jia for technical assistance by measuring the concentrations of phthalate metabolites.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names is for identification only and does not imply endorsement by the CDC, the Public Health Service, or the U.S. Department of Health and Human Services.

REFERENCES

- Adibi JJ, Whyatt RM, Williams PL, Calafat AM, Camann D, Herrick R, et al. 2008 Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. *Environ Health Perspect* 116:467–473. [PubMed: 18414628]
- Barkoski JDHB, Tancredi D, Boyd-Barr D, Elms W, Hertz-Picciotto I. In press. Variability of urinary pesticide metabolite concentrations during pregnancy in the marbles study. *Environ Res*
- Braun JM, Smith KW, Williams PL, Calafat AM, Berry K, Ehrlich S, et al. 2012 Variability of urinary phthalate metabolite and bisphenol a concentrations before and during pregnancy. *Environmental Health Perspectives* 120:739–745. [PubMed: 22262702]
- Calafat AM. 2016 Contemporary issues in exposure assessment using biomonitoring. *Current Epidemiology Reports* 3:145–153. [PubMed: 28884070]
- Cantonwine DE, Cordero JF, Rivera-Gonzalez LO, Del Toro LVA, Ferguson KK, Mukherjee B, et al. 2014 Urinary phthalate metabolite concentrations among pregnant women in northern puerto rico: Distribution, temporal variability, and predictors. *Environment International* 62:1–11. [PubMed: 24161445]
- CDC. 2017 Fourth national report on human exposure to environmental chemicals Atlanta, GA.
- Dodson RE, Nishioka M, Standley LJ, Perovich LJ, Brody JG, Rudel RA. 2012 Endocrine disruptors and asthma-associated chemicals in consumer products. *Environmental Health Perspectives* 120:935–943. [PubMed: 22398195]

- Enke U, Schleussner E, Palmke C, Seyfarth L, Koch HM. 2013 Phthalate exposure in pregnant women and newborns - the urinary metabolite excretion pattern differs distinctly. *International Journal of Hygiene and Environmental Health* 216:735–742. [PubMed: 23474103]
- Ferguson KK, McElrath TF, Ko YA, Mukherjee B, Meeker JD. 2014a Variability in urinary phthalate metabolite levels across pregnancy and sensitive windows of exposure for the risk of preterm birth. *Environment International* 70:118–124. [PubMed: 24934852]
- Ferguson KK, McElrath TF, Meeker JD. 2014b Environmental phthalate exposure and preterm birth. *Jama Pediatr* 168:61–67. [PubMed: 24247736]
- Fisher M, Arbuckle TE, Mallick R, LeBlanc A, Hauser R, Feeley M, et al. 2015 Bisphenol a and phthalate metabolite urinary concentrations: Daily and across pregnancy variability. *Journal of Exposure Science and Environmental Epidemiology* 25:231–239. [PubMed: 25248937]
- Fromme H, Bolte G, Koch HM, Angerer J, Boehmer S, Drexler H, et al. 2007 Occurrence and daily variation of phthalate metabolites in the urine of an adult population. *International Journal of Hygiene and Environmental Health* 210:21–33. [PubMed: 17182278]
- Hauser R, Meeker JD, Park S, Silva MJ, Calafat AM. 2004 Temporal variability of urinary phthalate metabolite levels in men of reproductive age. *Environ Health Perspect* 112:1734–1740. [PubMed: 15579421]
- Hauser R, Calafat AM. 2005 Phthalates and human health. *Occup Environ Med* 62:13. [PubMed: 15613603]
- Heffernan AL, Aylward LL, Toms LML, Sly PD, Macleod M, Mueller JF. 2014 Pooled biological specimens for human biomonitoring of environmental chemicals: Opportunities and limitations. *Journal of Exposure Science and Environmental Epidemiology* 24:225–232. [PubMed: 24192659]
- Hertz-Picciotto I, Schmidt RJ, Walker CK, Bennett DH, Oliver M, Wise K, Giulivi C, Puschner B, Thomas J, Lasalle JM, Pessah I, Van de Water J, Ozonoff S. A prospective study of environmental exposures and early biomarkers in Autism Spectrum Disorder: the MARBLES Study. *Environmental Health Perspectives* (in press).
- Heudorf U, Mersch-Sundermann V, Angerer E. 2007 Phthalates: Toxicology and exposure. *International Journal of Hygiene and Environmental Health* 210:623–634. [PubMed: 17889607]
- Hoppin JA, Brock JW, Davis BJ, Baird DD. 2002 Reproducibility of urinary phthalate metabolites in first morning urine samples. *Environmental Health Perspectives* 110:515–518. [PubMed: 12003755]
- Hornung RW, Reed LD. 1990 Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene* 5:46–51.
- Huang PC, Kuo PL, Chou YY, Lin SJ, Lee CC. 2009 Association between prenatal exposure to phthalates and the health of newborns. *Environment International* 35:14–20. [PubMed: 18640725]
- Koch HM, Calafat AM. 2009 Human body burdens of chemicals used in plastic manufacture. *Philosophical Transactions of the Royal Society B-Biological Sciences* 364:2063–2078.
- Koch HM, Lorber M, Christensen KLY, Palmke C, Koslitz S, Bruning T. 2013 Identifying sources of phthalate exposure with human biomonitoring: Results of a 48 h fasting study with urine collection and personal activity patterns. *Int J Hyg Envir Heal* 216:672–681.
- Lin S, Ku H-Y, Su P-H, Chen J-W, Huang P-C, Angerer J, et al. 2011 Phthalate exposure in pregnant women and their children in central taiwan. *Chemosphere* 82:947–955. [PubMed: 21075419]
- Meeker JD, Hu H, Cantonwine DE, Lamadrid-Figueroa H, Calafat AM, Ettinger AS, et al. 2009 Urinary phthalate metabolites in relation to preterm birth in mexico city. *Environmental Health Perspectives* 117:1587–1592. [PubMed: 20019910]
- Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. 2011 Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. *Pediatrics* 128:E488–E495. [PubMed: 21844053]
- Perrier F, Giorgis-Allemand L, Slama R, Philippat C. 2016 Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology* 27:378–388. [PubMed: 27035688]
- Preau JL, Wong LY, Silva MJ, Needham LL, Calafat AM. 2010 Variability over 1 week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among eight

- adults: An observational study. *Environmental Health Perspectives* 118:1748–1754. [PubMed: 20797930]
- Romano ME, Hawley NL, Eliot M, Calafat AM, Jayatilaka NK, Kelsey K, et al. 2017 Variability and predictors of urinary concentrations of organophosphate flame retardant metabolites among pregnant women in Rhode Island. *Environmental Health*:40.
- Rosenberg J, Fiserova-Bergerova V, Lowry LK. 1989 Biological monitoring iv: Measurements in urine. *Applied Industrial Hygiene* 4:F16–F21.
- Rosner B 2000 *Fundamentals of biostatistics* 5th ed. Pacific Grove, CA:Duxbury.
- Rudel RA, Gray JM, Engel CL, Rawsthorne TW, Dodson RE, Ackerman JM, et al. 2011 Food packaging and bisphenol a and bis(2-ethylhexyl) phthalate exposure: Findings from a dietary intervention. *Environmental Health Perspectives* 119:914–920. [PubMed: 21450549]
- Schettler T 2006 Human exposure to phthalates via consumer products. *Int J Androl* 29:134–139. [PubMed: 16466533]
- Silva MJ, Samandar E, Preau JL, Reidy JA, Needham LL, Calafat AM. 2007 Quantification of 22 phthalate metabolites in human urine. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 860:106–112.
- Varshavsky JR, Morello-Frosch R, Woodruff TJ, Zota AR. 2018 Dietary sources of cumulative phthalates exposure among the US general population in NHANES 2005–2014. *Environment International* 115:417–429. [PubMed: 29605141]
- Varshavsky JR, Morello-Frosch R, Woodruff TJ, Zota AR. In press. Dietary sources of cumulative phthalates exposure among the U.S. general population in NHANES 2005–2014. *Environ Int*
- Woodruff TJ, Zota AR, Schwartz JM. 2011 Environmental chemicals in pregnant women in the United States: NHANES 2003–2004. *Environmental Health Perspectives* 119:878–885. [PubMed: 21233055]
- Wormuth M, Scheringer M, Vollenweider M, Hungerbühler K. 2006 What are the sources of exposure to eight frequently used phthalic acid esters in Europeans? *Risk Anal* 26:803–824. [PubMed: 16834635]
- Ye XY, Wong LY, Bishop AM, Calafat AM. 2011 Variability of urinary concentrations of bisphenol a in spot samples, first morning voids, and 24-hour collections. *Environ Health Perspect* 119:983–988. [PubMed: 21406337]
- Zota AR, Calafat AM, Woodruff TJ. 2014 Temporal trends in phthalate exposures: Findings from the National Health and Nutrition Examination Survey, 2001–2010. *Environmental Health Perspectives* 122:235–241. [PubMed: 24425099]

Highlights

- ICCs were higher in pooled samples than in FMVs
- FMV/pool ratios tended to decrease with an increasing number of composites
- Pooling can improve characterization of individual's true exposure

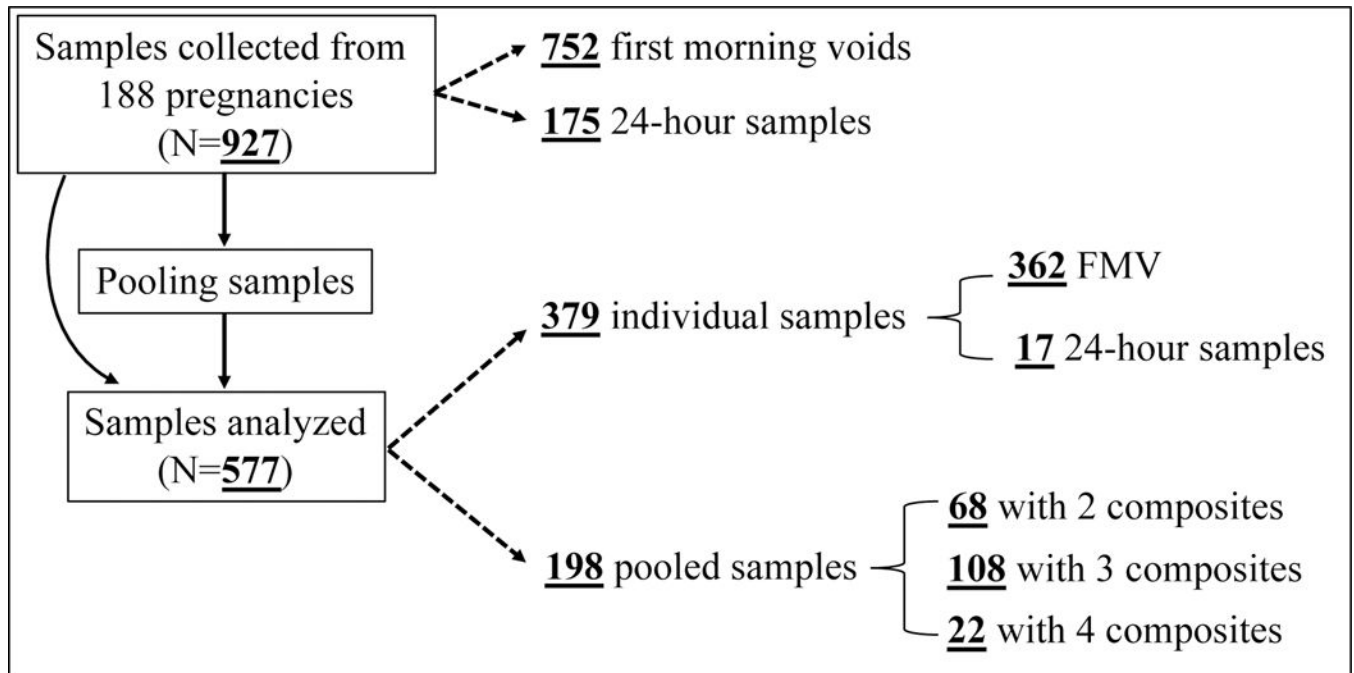


Figure 1.

Type and number of urine samples analyzed in this study from 188 pregnancies.

Theoretically, if all mothers were fully compliant with our sampling protocol (three FMVs and one 24-hour sample for each of the 2nd and 3rd trimesters), the total number of samples available for this present study should be 1,504. After selecting the first FMV as an individual sample and pooling the remaining samples, a total number of 752 samples should remain for analysis: 376 FMVs as an individual sample and 376 pools with 3 composites. Although all mothers were asked to collect four samples per trimester, many did not. In some cases, samples provided by mothers as the 2nd trimester sample were later determined as the 3rd trimester sample or vice versa. In other cases, mothers collected more samples than necessary. Thus, the number of samples included in each of the 198 pools varied based on the number of samples each woman collected.

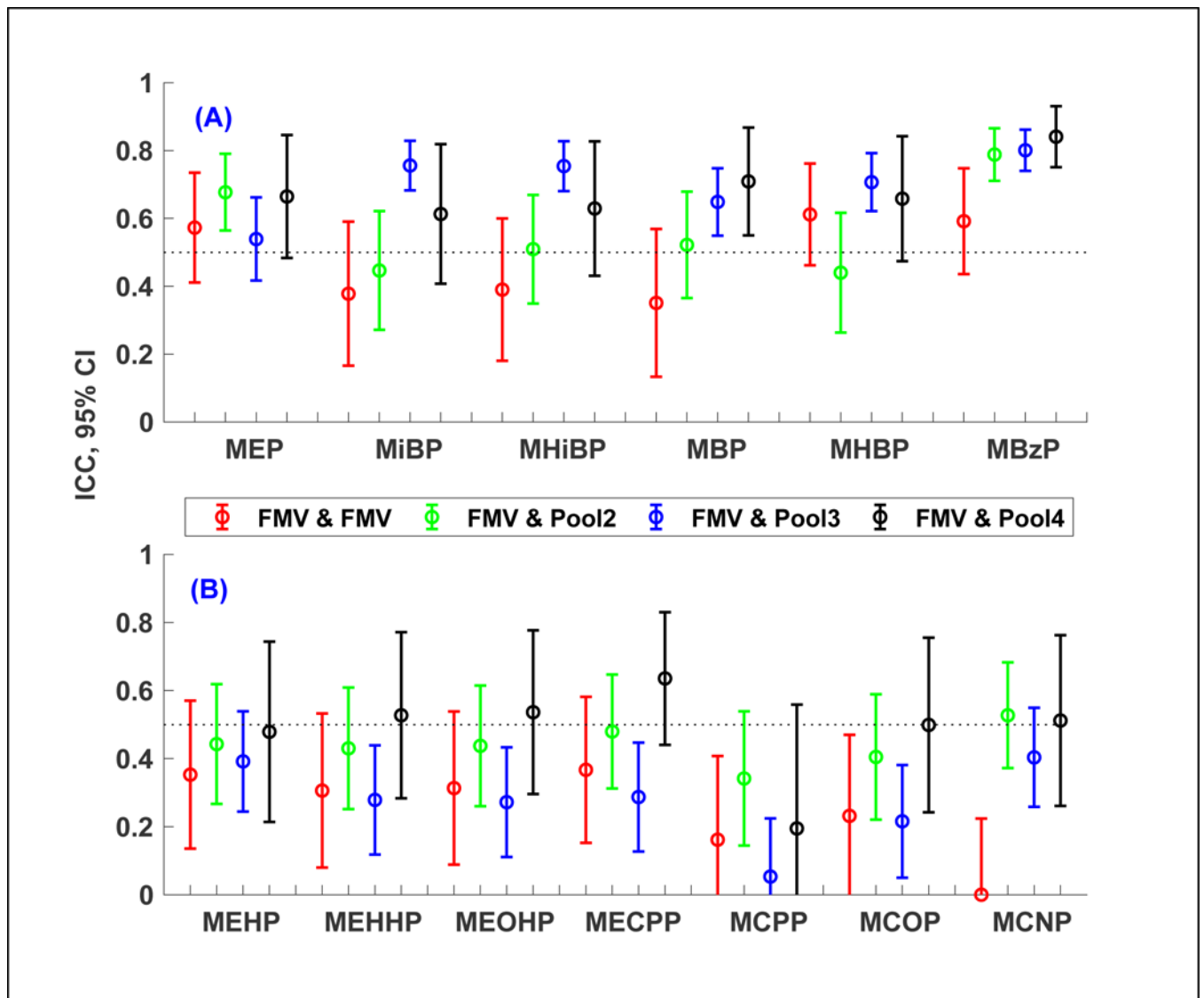


Figure 2.

Intra-class correlation coefficients (ICC) and 95% confidence interval (CI) of ln-transformed SG-corrected maternal urinary phthalate metabolite concentrations using samples collected within the same (2nd or 3rd) trimester from 162 pregnancies: (1) two first FMVs (first morning voids) in case when only FMVs were available each trimester ($n = 98$ from 41 mothers), (2). one FMV and one pooled sample with 2 composites ($n = 128$ from 62 mothers), (3) one FMV and one pooled sample with 3 composites ($n = 206$ from 85 mothers), and (4) one FMV and one pooled sample with 4 composites ($n = 44$ from 21 mothers). Two FMVs used in ICC calculations were collected approximately 1 week apart and the last sample in the pool and the first FMV were approximately 2, 3, or 4 weeks apart depending on the number of composites in the pool. Pooled samples with a different number (2, 3, or 4) of composite samples were indicated in legend as Pool2, Pool3, and Pool4, respectively. In the selected pooled samples, 59% of pools with 2 composites were comprised of at least one 24-hour sample, 91% of pools with 3 composites were comprised

of at least one 24-hour sample, and 82% of pools with 4 composites were comprised of at least one 24-hour sample. Refer to Table S4 for values of ICC and 95% CI.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

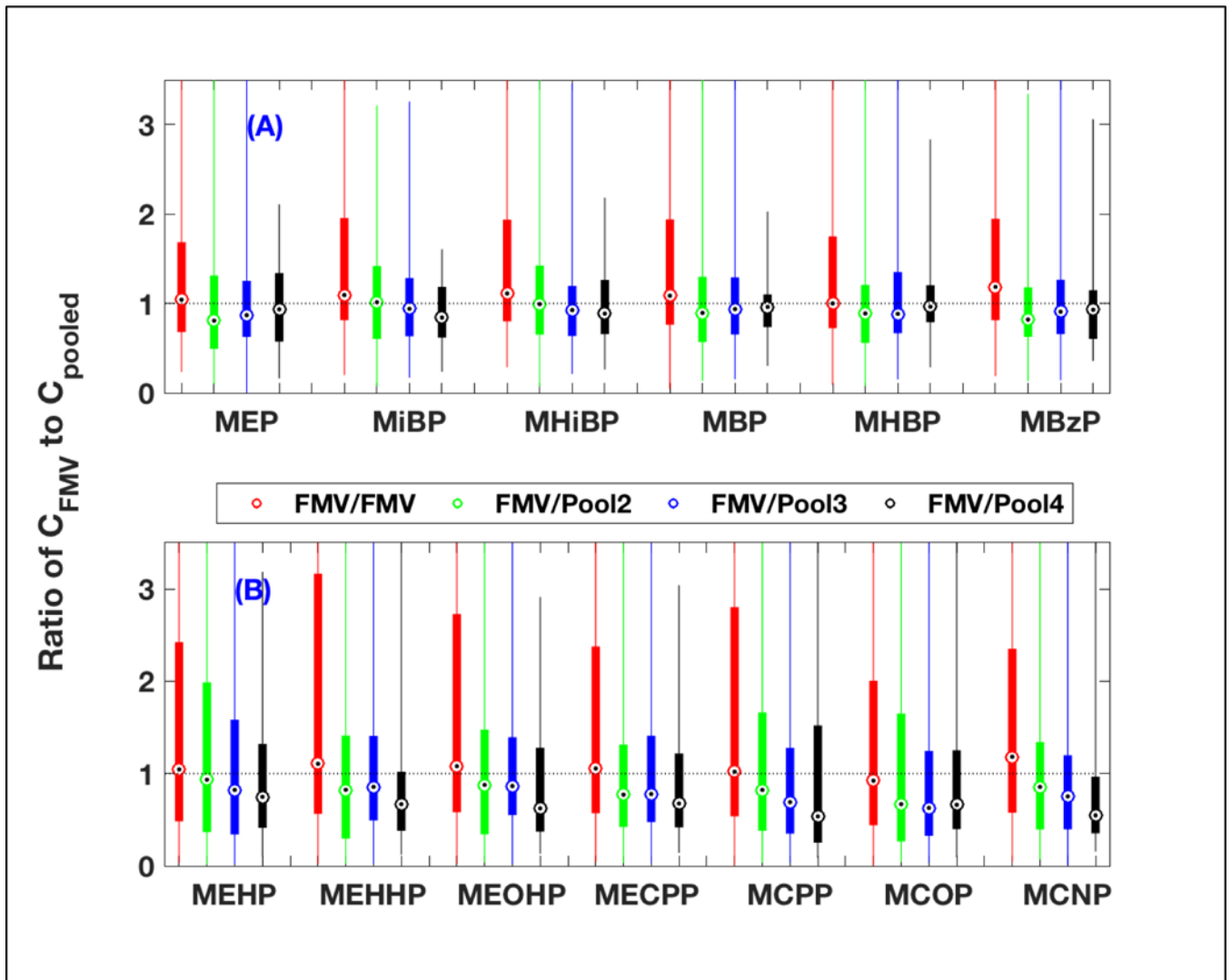


Figure 3.

The ratio of the urinary phthalate metabolite concentrations of a first morning void (FMV) to another FMV or an FMV to a pooled sample collected within the same trimester using the same dataset in Figure 2.

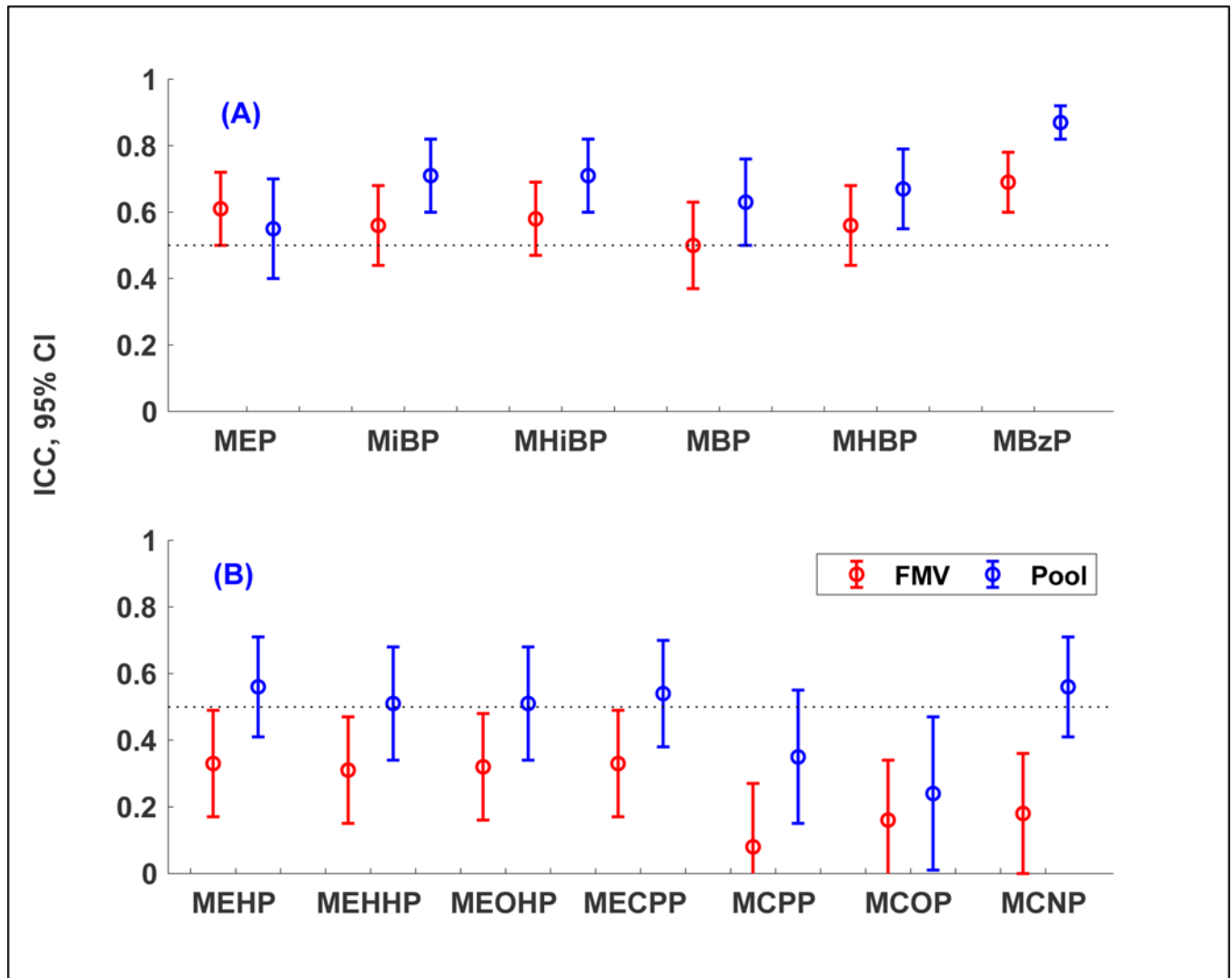


Figure 4.

Intraclass correlation coefficients (ICC) and 95% confidence interval (CI) of ln-transformed SG-corrected maternal urinary phthalate metabolite concentrations using samples collected during each of the 2nd and 3rd trimesters: (1) two first FMVs (first morning voids) ($n = 204$ from 102 mothers) and (2) two pooled samples ($n = 118$ from 59 mothers). Two samples of the same urine type (i.e., FMVs or pools) used in ICC calculations were collected approximately 3 months apart. The selected pooled samples in this figure were comprised of FMVs only ($n = 23$), or a varying number of FMVs plus one 24-hour sample ($n = 88$), or a varying number of FMVs plus two 24-hour samples ($n = 7$). Refer to Table S6 for values of ICC and 95% CI.

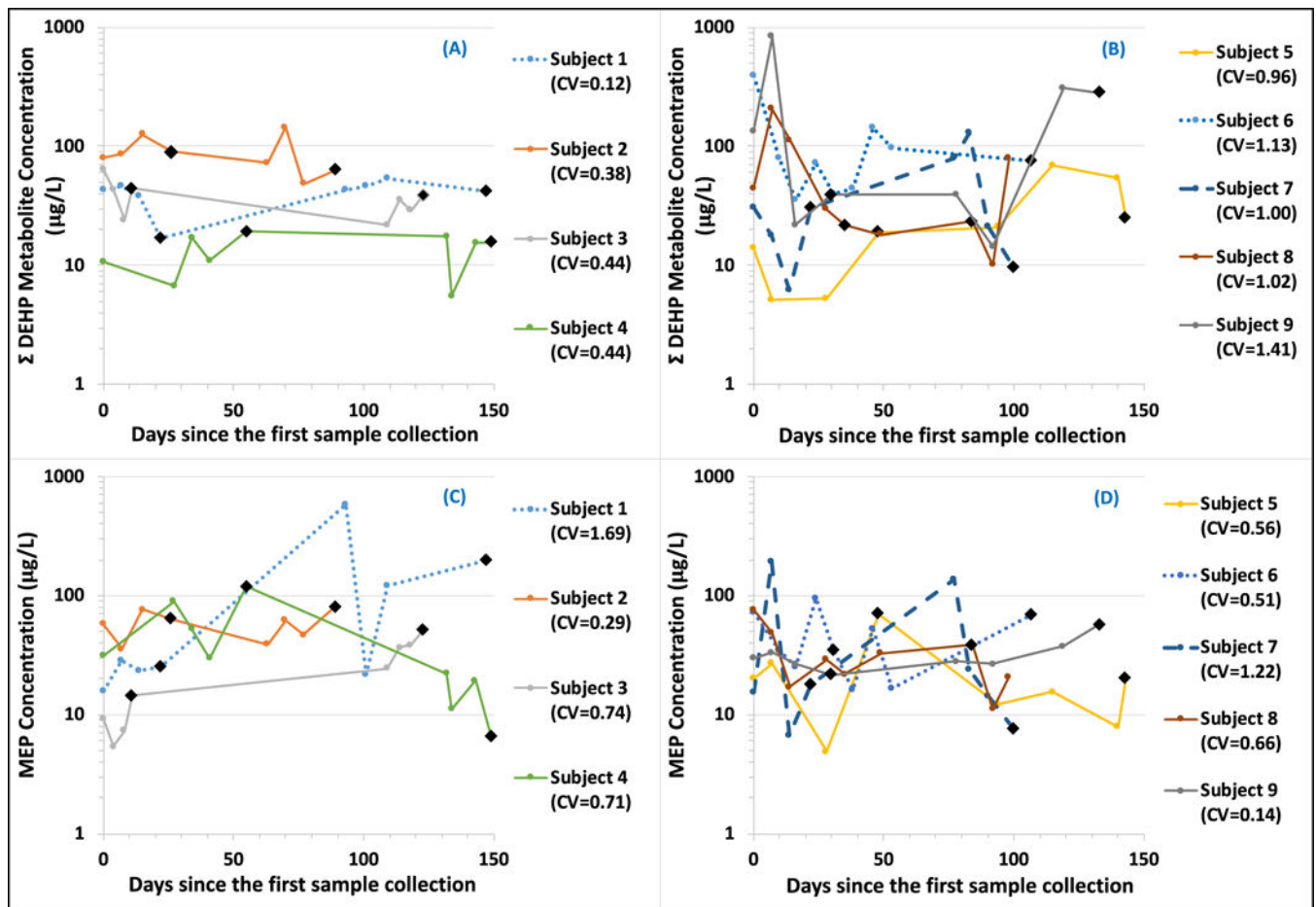


Figure 5.

Temporal variability of the sum of DEHP metabolites (Σ DEHP) and of MEP from longitudinal samples of 9 mothers, plotted versus the days since the first sample collection. First four or five samples were typically collected once a week during the 2nd trimester and the next four or five samples were typically collected once a week during the 3rd trimester. Circle points represent first morning voids (FMVs) and black diamond points 24-hour samples. Only FMVs were used to compute coefficient of variation (CV). The degree of within-subject variability varies not only between participants, but also by metabolites within participants.

Table 1.

Distribution of SG-corrected phthalate metabolite concentrations ($\mu\text{g/L}$) in 560 urine samples (FMVs and pooled samples only) collected from 188 unique pregnancies during the period from 2007 to 2014, and comparison of geometric mean (GM) with the United States general female population from 2011–2012 NHANES.

Metabolites	LOD ($\mu\text{g/L}$)	Percent > LOD	Percentile							Geometric mean (GM)		
			5 th	25 th	50 th	75 th	95 th	Maximum	This study (2007–2014)	This study (2011–2012) ^a	NHANES (2011–2012) ^b	
MEP	1.2	100%	5.6	13.4	22.6	48.9	184.0	3327.3	26.7	22.8	37.7	
MIBP	0.8	98%	2.0	4.7	7.3	12.0	23.9	99.8	7.2	7.6	5.5	
MHiBP	0.4	97%	0.8	1.6	2.6	4.2	9.0	38.8	2.6	2.7	NM	
MBP	0.4	99%	3.4	7.9	12.4	20.6	41.1	131.2	12.0	12.9	7.1	
MHBP	0.4	82%	<LOD	0.6	1.1	1.7	3.7	14.1	1.0	1.0	NM	
MBzP	0.3	99%	1.2	3.4	6.5	12.1	37.7	120.8	6.4	6.6	4.3	
MEHP	0.8	82%	<LOD	1.2	2.5	5.4	23.1	2833.8	2.7	1.7	1.2	
MEHHP	0.4	100%	2.5	6.5	11.2	22.7	98.4	6803.1	13.0	8.4	7.2	
MEOHP	0.2	100%	2.0	5.1	9.1	17.2	71.2	3646.2	10.0	6.7	4.7	
MECPP	0.4	100%	4.8	10.5	18.1	37.1	156.9	8852.3	21.8	13.7	11.8	
MCPP	0.4	92%	<LOD	1.0	1.7	3.4	14.9	203.6	1.9	2.0	2.6	
MNP	0.9	49%	<LOD	<LOD	0.9	2.0	8.6	123.0	1.1	1.1	NA	
MCOP	0.3	100%	2.9	6.5	13.2	33.3	155.6	968.0	16.1	17.7	17.1	
MCNP	0.2	100%	0.8	1.6	2.7	5.1	25.1	527.1	3.3	2.6	3.0	

Abbreviations - GM: geometric mean, NA: not available, NM: compound not measured, LOD: limit of detection, <LOD: below LOD

^aN = 63 first morning voids

^bNHANES: Females only between 2011 and 2012.