Journal of Exposure Science & Environmental Epidemiology (2018) 28:448–460
https://doi.org/10.1038/s41370-018-0050-0

ARTICLE

# A possible approach to improving the reproducibility of urinary concentrations of phthalate metabolites and phenols during pregnancy

Mahsa M. Yazdy[1,2] · Brent A. Coull[3,4] · Joseph C. Gardiner[5] · Andrea Aguiar[6] · Antonia M. Calafat[7] · Xiaoyun Ye[7] · Susan L. Schantz[6] · Susan A. Korrick[1,3]

Received: 10 July 2017 / Revised: 4 April 2018 / Accepted: 22 April 2018 / Published online: 17 July 2018


## Abstract

In epidemiologic studies, classifying episodic exposures to chemicals with short half-lives, such as phthalates and phenols, is challenging. We assessed whether accounting for sources of variability unrelated to exposure pathways would improve the reproducibility of urine concentrations of select phthalate metabolites and phenols. In 2011, a subset of pregnant women ($n = 19$) enrolled in a prospective study provided first morning urine samples every 3–4 weeks between 16 and 36 weeks gestation. At the time of collection, we identified potential contributors to variations in urinary concentrations: weight gain, gestational age, time slept, time since awoke, time since last food/drink, and time since last void. We estimated intraclass correlation coefficients (ICCs) among repeat urine concentrations with and without adjustment for sources of variability using a random intercept linear mixed model. Concentrations of monoethyl phthalate, butyl, and propyl parabens were the most reproducible (ICCs: 0.68, 0.56, and 0.56, respectively). However, adjustment for potential sources of variability unrelated to exposure pathways did not materially improve reproducibility nor the ability of a single sample to predict exposure based on average biomarker concentrations across pregnancy. Future studies should carefully consider the exposure timeframe and the reliability of using biomarker concentrations from a single time point to represent exposures over pregnancy.

**Keywords** Child exposure/health · Endocrine disruptors · Phenols · Phthalates · Pregnancy · Prenatal exposure

## Introduction

The potential for prenatal and childhood exposures to endocrine-disrupting chemicals (EDCs) to adversely impact early life development is a growing public health concern. Phenols and phthalates are two classes of EDCs that are commonly used in a wide range of consumer and personal care products thereby resulting in pervasive human exposure as shown by regular detection of EDC biomarkers in urine [1, 2]. A growing research effort is focused on assessing the potential contribution of phenols and phthalates to adverse neurobehavioral and reproductive development in childhood [3].

Bisphenol A (BPA) is a widely studied EDC for which human exposure occurs largely through diet due to leaching from food packaging and polycarbonate containers, though transdermal absorption (e.g., contact with thermal receipts) and inhalation (e.g., cigarette filters) are also possible [1, 4].

**Electronic supplementary material** The online version of this article (https://doi.org/10.1038/s41370-018-0050-0) contains supplementary material, which is available to authorized users.

✉ Susan A. Korrick
susan.korrick@channing.harvard.edu

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

2 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

3 Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

4 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

5 Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA

6 Department of Comparative Biosciences and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

7 Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

Other commonly occurring phenols include triclosan and parabens both of which have antimicrobial properties resulting in their use in personal care products (e.g., soaps, toothpastes, deodorants) and/or as preservatives in cosmetics, pharmaceuticals, and foods [1, 5]. 2,4-Dichlorophenol is predominantly found in herbicides but can also be formed as a by-product of manufacturing chlorinated chemicals [6]. 1,4-Dichlorobenzene, which metabolizes to 2,5-dichlorophenol, is used in moth balls and toilet deodorants [6]. Another phenolic compound, benzophenone-3 is used in sunscreen because of its capacity to block ultraviolet radiation [1]. Despite increasing use of substitutes for BPA and other common phenols in consumer products, these phenols are still detectable in human urine. Although typically short-lived in the body, a number of phenols, including benzophenone-3 and its metabolites, have been detected in lipid tissues supporting the possibility of bioaccumulation [7, 8].

Phthalates are a family of structurally related chemicals used commonly in many consumer and personal care products. For example, diethyl phthalate (DEP) is largely used in personal care products with perfume or fragrance while di(2-ethylhexyl) phthalate (DEHP) is used to produce flexible plastics (e.g., polyvinyl chloride) for a variety of household, garden, and medical products (e.g., intravenous tubing) [1].

The rapid growth and development that occurs prenatally can be particularly sensitive to perturbation by exposure to EDCs, such as certain phthalates and phenols [2]. To date, urinary concentrations of phenol and phthalate metabolites are the best exposure biomarkers [1]. But these EDCs have short half-lives and individual exposures may be highly variable over the course of a day and from day to day, making human exposure assessment challenging. Consistent with their short half-life and the episodic nature of the exposure, repeat urinary concentrations of phenol or phthalate biomarkers have poor reproducibility over time. In addition, for some biomarkers, reproducibility is even poorer in pregnant than in non-pregnant states [9–12]. As a result, there is uncertainty regarding how well urinary biomarker concentrations from a limited number of urine specimens (often only one) reflect exposure that is relevant to the risk of adverse health outcomes, including measures of child development. However, because of logistical and cost constraints, most epidemiologic studies of phenol or phthalate exposures rely on a limited number of urine samples per participant.

Research that would allow investigators to optimize the utility of urinary measures for exposure assessment in epidemiologic studies is needed. The goal of the current study was to assess whether there are easily ascertained sources of variability in urinary concentrations of phthalate metabolites and phenols during pregnancy that are unrelated to exposure pathways and that, when accounted for, would improve the reproducibility of the measurements.

## Materials and methods

### Study population

The formative (pilot) phase of the Illinois Kids Development Study (I-KIDS) enrolled 181 pregnant women between 2010 and 2012. I-KIDS is a prospective cohort study of the relation of prenatal exposures to phenols (or their precursors) and phthalates with fetal growth and sexual development as well as with subsequent infant and child cognition. The formative phase of I-KIDS ended in 2012 and an expanded version of the study is in progress. Women learned about the formative study at their first prenatal visit to a local obstetrics clinic in Urbana, IL and completed a reply card indicating their potential interest in participation. Of the 1280 women who completed a reply card, 512 (40%) indicated interest in the study. Of these 512, 400 (78%) were reached in a follow-up phone call and 224 (56%) were identified as eligible, 181 (81%) of whom enrolled in the study. After enrollment, 24 (13%) women became ineligible or withdrew from the study during pregnancy and an additional 8 (4%) became ineligible or withdrew from the study after delivery, resulting in 149 women in the final cohort. Eligible women were 18–40 years of age, fluent in English, not carrying multiples, and not taking prescription medication for a chronic health condition; they also resided within a 45-mile radius of the study clinic and planned to stay in the hospital for 48 h after delivery and in the area for at least 1 year after the birth of their infant.

The study protocol was reviewed and approved by the Human Subjects Committees of the University of Illinois at Urbana-Champaign and Carle Foundation Hospital, Urbana, IL. Written informed consent was obtained from participants before study assessments were performed. The analysis of blinded urine specimens by the Centers for Disease Control and Prevention (CDC) laboratory was determined not to constitute engagement in human subjects research.

### Urine sample collection

Women were enrolled in their second trimester of pregnancy, between 16 and 19 weeks gestation. They provided a first morning urine specimen at enrollment and a second one at the end of the third trimester, between 34 and 38 weeks gestation. For the current analysis, the 30 women living closest to the study clinic were asked to provide more frequent (every 3–4 weeks) first morning urine samples during pregnancy; 19 women agreed to do so. Each of the 19 women in this

sub-study collected a total of 6 urine samples between 16 and 36 weeks gestation.

Once every 3–4 weeks for the last 6 months of pregnancy (at 16–18, 20–22, 23–26, 27–30, 32–33, and 35–36 weeks gestation), each of the 19 woman collected a first morning urine sample using a polypropylene urine collection cup with a high-density polyethylene (HDPE) lid (Thermo Scientific Nalgene). Samples were kept refrigerated until processing, within approximately 24 h of collection. For processing, samples were warmed to room temperature, mixed on a vortex mixer, and had specific gravity measured using a refractometer (TS400; Reichert). The urine was then aliquoted with disposable polyethylene transfer pipets (Fisherbrand) into polypropylene vials with HDPE lids (Thermo Scientific Nalgene) and stored at −20 °C. Field blanks of purified water (Fisher Chemical W5SK-1 HPLC grade, submicron filtered) were collected periodically, transported to the lab, processed, and stored using the same procedures described for urine. Frozen urine samples and field blanks were shipped on dry ice by overnight courier to the CDC (Atlanta, GA).

## Covariate assessment

At enrollment, study participants completed a questionnaire to provide information on demographics, height, pre-pregnancy weight, occupation, lifestyle, reproductive and medical history, and exposure history. In addition, for the 24 h before each urine collection, participants completed a diary recording use of products potentially containing phenols (or their precursors) or phthalates (e.g., personal care products, pre-packaged foods). The women also recorded the time of their last meal or drink and last void prior to collecting the urine samples. Finally, trained study staff reviewed medical records from prenatal visits to collect pregnancy weight. To calculate maternal weight gain at each urine collection, we identified the maternal weight in the medical record that was measured closest (within 14 days) to the urine collection date and subtracted it from the pre-pregnancy weight.

## Urine phthalate metabolites and phenol measurements

Total (free plus conjugated) concentrations of eight phenols (BPA, triclosan, 2,4-dichlorophenol, 2,5-dichlorophenol, benzophenone-3 as well as butyl paraben (B-paraben), methyl paraben (M-paraben), and propyl paraben (P-paraben)) and 11 phthalate metabolites were quantified at CDC by online solid phase extraction coupled with high-performance liquid chromatography–isotope dilution–tandem mass spectrometry [13–16]. The CDC laboratory methods have excellent sensitivity and reproducibility for these urine analyses with coefficients of variation (CVs) ranging from 2.7% to 15% [15–17].

For this analysis, we examined the total concentrations of the eight phenols and the metabolites of two commonly used phthalates that accounted for the majority of phthalate urine biomarkers in our population: (1) monoethyl phthalate (MEP), the metabolite of DEP; and (2) the micromolar sum (μmol/L) of four metabolites of DEHP: mono (2-ethylhexyl) phthalate (MEHP), mono (2-ethyl-5-hydroxyhexyl) phthalate (MEHHP), mono (2-ethyl-5-oxohexyl) phthalate (MEOHP), and mono (2-ethyl-5-carboxypentyl) phthalate (MECPP). We calculated the molar sum of thefour DEHP metabolites (∑DEHP) by dividing the concentration of each metabolite by its molar mass and summing the resulting measures. In the 149 women participating in the full formative center study, MEP accounted for approximately 38% and ∑DEHP metabolites accounted for 18% of the phthalate metabolites detected in urine.

## Statistical analysis

Concentrations of all urine biomarkers (including results below the limit of detection for which a signal was detected) were used as reported in the analyses after natural log transformation. Only two biomarkers had measured concentrations of zero (e.g., no signal detected)—triclosan and B-paraben (12 urine samples from 8 women and 31 urine samples from 10 women, respectively). Prior to in-transformation, these results were assigned 0.001 μg/L, which was 1% of the lowest non-zero measured concentration. Non-reportable results that did not meet pre-established laboratory quality control criteria were set to missing. If >10% of measurements for a specific biomarker were missing, the biomarker was not included in the statistical analyses. To account for urine dilution, biomarker concentrations were adjusted for specific gravity using the following formula: $B_{SG} = B \times [(1.024 - 1)/(SG - 1)]$, where $B_{SG}$ is the specific gravity adjusted biomarker concentration, $B$ is the measured biomarker value, and SG is the urine specific gravity [18].

We identified covariates based on *a priori* considerations related to the short half-life of the biomarkers and the potential for time-varying activities (e.g., time since last void) to impact urinary concentrations. We did not consider potential sources of exposure (e.g., personal care product use) as covariates in this analysis because our goal was to improve reproducibility over time by removing within-person variation in urine measures that arise from differences in conditions at the time of urine collection unrelated to exposure itself. Our goal was not to predict biomarker concentrations nor to identify all sources of within person variability. For example, we wanted to retain variability related to suspected or known exposure pathways, as this is

**Table 1** Characteristics of I-KIDS mothers who provided six pregnancy urine samples ($n = 19$) compared to the full cohort participating in the I-KIDS formative study ($n = 149$)

| Maternal characteristics | Full cohort ($n = 149$) | Subgroup analysis ($n = 19$) | $P$ value[a] |
|---|---|---|---|
| Age (years); mean ± SD and $n$ (%) | 29.5 ± 4.1 | 29.5 ± 3.5 | 0.99 |
| <20 | 3 (2.0) | 0 (0.0) | |
| 20–24 | 15 (10.1) | 1 (5.3) | |
| 25–29 | 59 (39.6) | 7 (36.8) | |
| 30–34 | 55 (36.9) | 11 (57.9) | |
| ≥35 | 17 (11.4) | 0 (0.0) | |
| Race/ ethnicity; $n$ (%) | | | |
| Non-Hispanic White | 125 (83.9) | 17 (89.5) | 0.53 |
| Non-Hispanic Black | 8 (5.4) | 1 (5.3) | |
| Hispanic | 3 (2.0) | 1 (5.3) | |
| Asian | 7 (4.7) | 0 (0.0) | |
| Multiracial or other race/ethnicity | 6 (4.0) | 0 (0.0) | |
| Education; $n$ (%) | | | |
| ≤12 years | 6 (4.0) | 0 (0.0) | 0.82 |
| Some college | 35 (23.5) | 5 (26.3) | |
| College graduate | 108 (72.5) | 14 (73.7) | |
| Income; $n$ (%) | | | |
| <$30,000 | 24 (16.1) | 4 (21.1) | 0.72 |
| $30,000–$59,999 | 33 (22.2) | 4 (21.1) | |
| $60,000–$89,999 | 57 (38.3) | 7 (36.8) | |
| ≥$90,000 | 32 (21.5) | 4 (21.1) | |
| Missing | 3 (2.0) | 0 (0.0) | |
| Marital status; $n$ (%) | | | |
| Married | 127 (85.2) | 18 (94.7) | 0.41 |
| Co-habitating as married | 14 (9.4) | 1 (5.3) | |
| Single | 8 (5.4) | 0 (0.0) | |
| Number of previous live born children; $n$ (%) | | | |
| 0 | 54 (36.2) | 6 (31.6) | 0.50 |
| 1 | 58 (38.9) | 7 (36.8) | |
| ≥2 | 37 (24.8) | 6 (31.6) | |
| Smoking during pregnancy; $n$ (%) | | | |
| Yes | 5 (3.4) | 1 (5.3) | 0.62 |
| No | 144 (96.6) | 18 (94.7) | |
| Pre-pregnancy BMI (kg/m$^2$); $n$ (%) | | | |
| Underweight (<18.5) | 1 (0.7) | 0 (0.0) | 0.10 |
| Normal (18.5–24.9) | 93 (62.4) | 15 (79.0) | |
| Overweight (25–29.9) | 30 (20.1) | 4 (21.1) | |
| Obese (≥30) | 23 (15.4) | 0 (0.0) | |
| Missing | 2 (1.3) | 0 (0.0) | |
| Gestational age at birth (weeks); mean ± SD | 39.3 ± 1.0 | 39.5 ± 1.3 | 0.58 |

*BMI* body mass index, *SD* standard deviations

[a]$P$ value comparing $n = 19$ in this analysis to the $n = 130$ excluded from this analysis. Wilcoxon–Mann–Whitney test was used for ordinal variables and chi-square test for the other categorical variables

central to the utility of biomarker concentrations for health effect studies. Instead, we focused on individual time-varying behaviors that demonstrate substantial within-person variability over time but are not related to differences in exposure itself. This is the primary reason we did not include covariates that are constant across a given person's urine collections (e.g., race, age, pre-pregnancy body mass index (BMI)). The chosen covariates included pregnancy weight gain (kg), gestational age (weeks), time since awoke (hours), time slept (hours), time since last food or drink (hours), and time since last void (hours) ascertained at the time of each urine collection. Covariate data were relatively complete—of the 684 time-varying covariate measures (19 women observed 6 times with 6 time-varying covariates per observation), a total of 35 values were missing. The few covariates with missing values were imputed using multiple imputation methods; in SAS [19] we created 10 imputed datasets using Proc MI.

We used multivariable linear mixed models to model the relationship of the selected covariates with urinary concentrations of phenols or phthalate metabolites. We assessed the reproducibility of repeat urinary concentrations by estimating unadjusted intraclass correlation coefficients (ICCs) and then re-estimating the ICCs after adjustment for potential sources of variability (unrelated to exposure pathways) using our linear mixed models. ICCs and corresponding 95% confidence intervals (CIs) were estimated from a random intercept linear mixed model. The ICC is the ratio of the between-subject variability and the sum of the between- and within-subject variability [20]; an ICC of 0 indicates no reproducibility within a subject and 1 indicates perfect reproducibility.

We also assessed how well concentrations of phthalate metabolites or phenols in a single urine sample predicted average urine concentrations over the entire pregnancy. First, we obtained studentized residuals from unadjusted and adjusted linear mixed models predicting urine biomarker concentrations for each of the six collection times. The residuals were the differences between the observed concentration in each woman individually at a given time point and the predicted value for each woman at each time point. We then used these residuals to assess how well biomarker concentrations in a single sample predict the overall pregnancy mean concentration (which was calculated as the mean of all six urine concentration residuals). We refer to the latter as our "gold standard" measure, assuming that a biomarker's average concentration across pregnancy is a more reliable measure of pregnancy exposure than the concentration at a single time point. This assessment included calculation of sensitivity, specificity, and percentage of correct classification using adjusted and unadjusted residuals. We grouped the distribution of residuals for each biomarker concentration at each collection

**Table 2** Median (range) of pregnancy urinary concentrations of phenols and phthalate metabolites among I-KIDS participants providing six pregnancy urine samples ($n = 19$ women)[a]

| Biomarker [LOD] | Pregnancy urine collection time in I-KIDS | | | | | | | | | | | | | Previous literature in pregnant women (range of median values, unadjusted for urine dilution) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 16–18 weeks | | 20–22 weeks | | 23–26 weeks | | 27–30 weeks | | 32–33 weeks | | 35–36 weeks | | Total[c], median (range) | |
| | *N* < LOD[b]/ NR | Median (range) | *N* < LOD[b]/ NR | Median (range) | *N* < LOD[b]/ NR | Median (range) | *N* < LOD[b]/ NR | Median (range) | *N* < LOD[b]/ NR | Median (range) | *N* < LOD[b]/ NR | Median (range) | | |
| Bisphenol A (μg/L) [0.4] | 1/0 | 1.3 (0.2–6.4) | 1/0 | 1.2 (0.3–7.0) | 1/0 | 1.3 (0.4–4.4) | 0/0 | 1.6 (0.6–8.4) | 0/0 | 1.2 (0.4–8.6) | 1/0 | 1.4 (0.2–6.8) | 1.5 (0.7–4.8) | 0.6–2.7 [9, 10, 28, 30–32, 34, 35, 43] |
| 2,4-Dichlorophenol (μg/L) [0.2] | 0/0 | 0.9 (0.5–6.0) | 0/0 | 0.8 (0.4–2.9) | 0/0 | 0.9 (0.4–3.1) | 0/0 | 0.7 (0.4–2.4) | 0/0 | 0.7 (0.2–2.5) | 1/0 | 0.8 (0.2–2.8) | 0.9 (0.6–3.1) | 0.5–2.1 [9, 10, 28–32] |
| 2,5-Dichlorophenol (μg/L) [0.2] | 0/0 | 0.7 (0.3–5.6) | 0/0 | 1.1 (0.3–3.1) | 0/0 | 1.6 (0.2–6.6) | 0/0 | 1.3 (0.4–10.8) | 0/0 | 1.4 (0.5–10.7) | 1/0 | 1.4 (0.2–6.9) | 1.5 (0.6–4.0) | 2.7–53 [9, 10, 28–32] |
| Benzophenone-3 (μg/L) [0.4] | 0/1 | 80.1 (3.0–809) | 0/1 | 49.0 (4.0–849) | 0/3 | 46.3 (1.6–861) | 0/3 | 30.8 (7.3–995) | 0/5 | 72.9 (4.1–883) | 0/3 | 102.0 (4.9–904) | 91.5 (19.1–762) | 1.7–77 [9, 10, 28, 30–32, 35] |
| Butyl paraben (μg/L) [0.2] | 10/0 | 0.2 (0–22.6) | 8/0 | 0.3 (0–3.1) | 9/0 | 0.2 (0–24.4) | 10/0 | 0.1 (0–15.6) | 12/0 | 0.1 (0–170) | 12/0 | 0.1 (0.001–5.8) | 0.4 (0–39.4) | 0.4–2.4 [9, 10, 28, 31] |
| Methyl paraben (μg/L) [1.0] | 0/1 | 108.5 (7.4–566) | 0/0 | 65.1 (7.5–233) | 0/0 | 85.0 (3.5–355) | 0/0 | 83.2 (8.7–880) | 0/1 | 129.5 (10.4–833) | 0/0 | 76.6 (2.2–445) | 111.1 (14.8–318) | 84.7–272 [9, 10, 28, 30, 31] |
| Propyl paraben (μg/L) [0.2] | 0/0 | 16.6 (0.3–279) | 0/0 | 16.3 (0.8–115) | 0/0 | 21.8 (0.3–84.3) | 1/0 | 12.0 (0.1–97.1) | 0/0 | 24.3 (0.5–849) | 0/0 | 6.2 (0.4–124) | 31.0 (0.4–212) | 12.5–45.6 [9, 10, 28, 30, 31] |
| Triclosan (μg/L) [2.3] | 7/0 | 5.3 (0–309) | 10/0 | 0.9 (0–178) | 4/0 | 6.1 (0–161) | 3/0 | 4.6 (1.1–286) | 3/11 | 19.5 (0.1–124) | 3/0 | 9.4 (1.2–187) | 14.8 (1.1–203) | <LOD–26.2 [9, 10, 28, 30–33, 35, 43] |
| Monoethyl phthalate (μg/L) [0.6] | 0/0 | 31.5 (5.7–709) | 0/0 | 26.6 (2.7–870) | 0/0 | 21.2 (4.9–349) | 0/0 | 32.3 (8.4–258) | 0/0 | 23.8 (3.1–202) | 0/0 | 30.8 (5.2–103) | 33.2 (5.8–356) | 21.5–380 [23, 28, 31, 32, 34–41] |
| ∑DEHP metabolites (μmol/L)[d] | 3[e]/0 | 0.1 (0.02–4.2) | 4[e]/0 | 0.1 (0.02–1.3) | 4[e]/0 | 0.1 (0.01–6.3) | 2[e]/0 | 0.1 (0.03–0.3) | 0/0 | 0.1 (0.03–0.9) | 7[e]/0 | 0.1 (0.02–1.1) | 0.1 (0.03–1.8) | 0.072–0.31[f] [32, 37, 39, 42] |

*LOD* limit of detection, *NR* no result, *∑DEHP metabolites* sum of metabolites of di(2-ethylhexyl) phthalate

[a]Values are not adjusted for urinary dilution (e.g., they are not adjusted for specific gravity)

[b]Measured concentrations <LOD for which a signal was detected, as well as those for which no signal was detected (i.e., concentration = 0), are reported in summary statistics

[c]Total = the mean biomarker concentration for each woman's 6 urine samples

[d]LODs (in μg/L) for the 4 DEHP metabolites were 0.5 for mono-2-ethylhexyl phthalate (MEHP) and 0.2 for mono-2-ethyl-5-carboxypentyl phthalate, mono-2-ethyl-5-hydroxyhexyl phthalate, and mono-2-ethyl-5-oxohexyl phthalate

[e]For ∑DEHP, the detection frequency of all metabolites, except MEHP, was 100%. Where MEHP was <LOD, ∑DEHP was assigned as <LOD

[f]Includes mean in addition to median values

time into tertiles and similarly grouped the distribution of gold standard residuals into tertiles. We defined "high exposure" as the top concentration tertile and "low exposure" as the two lowest tertiles. We then estimated how well a single urine identified a woman as having high overall pregnancy exposure by calculating the sensitivity, specificity, and percentage of correctly classified comparing each urine collection time with the gold standard. Finally, we compared results from unadjusted and adjusted analyses to determine whether accounting for potential sources of variability unrelated to exposure pathways improved the reproducibility, sensitivity, specificity, or correct classification of urinary concentrations of phthalate metabolites or phenols during pregnancy.

All analyses were conducted in each of the ten datasets for which missing covariate data were imputed. Results did not differ substantially among the ten datasets for the reproducibility analysis; therefore, we present ICCs from the first imputed dataset for this analysis. The results did differ across the ten datasets for the sensitivity and specificity analysis; therefore, we present the range of sensitivity, specificity, and percentage of correct classification that were identified across the ten imputed datasets.

All analyses were performed using the Statistical Analysis Software version 9.4 (SAS Institute Inc., Cary, NC).

## Results

### Study population

Participants in this sub-study ($n = 19$) were similar to those in the full formative study ($n = 149$). Consistent with the demographics of Urbana-Champaign which is a university community, they were mostly middle class (79% had household income ≥$30,000; vs. 82% in the full study), 90% were Caucasian (vs. 84% in the main study), and 74% were college educated (vs. 73% in the full study). All 19 women gave birth to full-term infants (mean 39.5 weeks) (Table 1). Mean age at enrollment was 29.5 years (range 23–34), most women (68%) were multiparous, and 37% reported ever smoking but only 1 (5%) smoked during pregnancy. For this sub-study, the majority of women (79%) reported a normal pre-pregnancy BMI (18.5–24.9 kg/m$^2$) and the average weight gain in mid-pregnancy (20–22 weeks) and late pregnancy (2–36 weeks) was 5.9 kg (standard deviation (SD): 4.9 kg) and 12.9 kg (SD: 6.0 kg), respectively. Women reported an average of 8.4 (SD: 0.8) hours of sleep the night before their first morning urine collection. Prior to the urine collection, the average number of hours since a woman voided or consumed food/drinks was 6.2 (SD: 1.6) and 8.3 (SD: 2.4), respectively.

**Table 3** Univariate determinants of urinary concentrations of phenols and phthalate metabolites (SG adjusted) during pregnancy, percent change (95% CI)

| Determinants | Bisphenol A ($n = 114$) | 2,4-dichlorophenol ($n = 114$) | 2,5-dichlorophenol ($n = 114$) | Butyl paraben ($n = 114$) | Methyl paraben ($n = 112$) | Propyl paraben ($n = 114$) | Monoethyl phthalate ($n = 114$) | ΣDEHP metabolites ($n = 114$) |
|---|---|---|---|---|---|---|---|---|
| | % change (95% CI) | % change (95% CI) | % change (95% CI) | % change (95% CI) | % change (95% CI) | % change (95% CI) | % change (95% CI) | % change (95% CI) |
| Gestational age (weeks) | 1.3 (−0.3, 2.9) | −2.5 (−4.4, −0.7) | 3.0 (0.7, 5.4) | −5.2 (−13.4, 3.0) | 0.1 (−2.3, 2.5) | −2.4 (−6.3, 1.5) | −0.6 (−2.9, 1.7) | 0.5 (−1.8, 2.7) |
| Weight gain (kg) | 2.5 (0.4, 4.7) | −1.6 (−3.9, 0.7) | 2.0 (−1.2, 5.2) | 0.0 (−11.1, 11.1) | 0.8 (−3.0, 4.5) | −0.7 (−6.8, 5.3) | 1.1 (−2.0, 4.2) | −0.1 (−2.9, 2.7) |
| Time since awoke (h) | −5.6 (−29.8, 18.6) | −25.2 (−40.3, −10.0) | −18.0 (−41.4, 5.5) | −38.6 (−112.8, 35.5) | −7.2 (−37.7, 23.4) | 0.7 (−46.8, 48.1) | −0.2 (−16.4, 16.0) | 13.1 (−11.8, 38.0) |
| Time slept (h) | 8.9 (−2.1, 19.8) | 13.6 (5.7, 21.5) | 6.8 (−4.4, 18.0) | 7.4 (−29.1, 43.9) | −2.9 (−18.4, 12.6) | −6.7 (−28.2, 14.7) | −0.8 (−11.5, 10.0) | −3.3 (−13.6, 7.0) |
| Time since last food/drink (h) | 2.8 (−0.5, 6.0) | −1.0 (−3.4, 1.4) | 0.8 (−2.2, 3.7) | −6.7 (−17.1, 3.6) | −2.4 (−7.2, 2.4) | −4.5 (−10.7, 1.6) | 0.2 (−3.0, 3.3) | 1.4 (−1.6, 4.4) |
| Time since last void (h) | 8.6 (4.0, 13.2) | 3.6 (0.3, 6.9) | 4.1 (−0.4, 8.6) | 0.0 (−15.2, 15.2) | 3.8 (−2.7, 10.2) | 8.5 (−0.5, 17.6) | 2.9 (−1.8, 7.5) | 4.9 (0.4, 9.5) |

Maximum number of observations is 114 (19 women × 6 samples each)

*ΣDEHP metabolites* sum of metabolites of di(2-ethylhexyl) phthalate

**Table 4** Intraclass correlation coefficients for repeat urinary concentrations of phenols and phthalate metabolites (SG adjusted) before and after adjustment for potential sources of variability unrelated to exposure pathways

| Biomarker | Unadjusted ICC (95% CI) | Adjusted[a] ICC (95% CI) | Previous literature in pregnant women ICCs adjusted for urinary concentration[b] |
|---|---|---|---|
| Bisphenol A | 0.24 (0.10, 0.49) | 0.20 (0.06, 0.47) | 0.04–0.31 [4, 9–12, 25, 46, 47] |
| 2,4-Dichlorophenol | 0.43 (0.24, 0.64) | 0.46 (0.25, 0.68) | 0.38–0.60 [9, 10] |
| 2,5-Dichlorophenol | 0.34 (0.16, 0.57) | 0.39 (0.20, 0.62) | 0.49–0.61 [9, 10] |
| Butyl paraben | 0.56 (0.36, 0.74) | 0.59 (0.38, 0.76) | 0.38–0.56 [9, 10, 12, 48] |
| Methyl paraben | 0.44 (0.25, 0.65) | 0.42 (0.22, 0.64) | 0.24–0.61 [9, 10, 12, 48] |
| Propyl paraben | 0.56 (0.36, 0.74) | 0.54 (0.34, 0.73) | 0.32–0.62 [9, 10, 12, 48] |
| Monoethyl phthalate | 0.68 (0.50, 0.82) | 0.66 (0.46, 0.81) | 0.21–0.50 [11, 23–26] |
| ∑DEHP metabolites | 0.32 (0.15, 0.55) | 0.27 (0.10, 0.57) | 0.08–0.31 [11, 23–26] |

*∑DEHP metabolites* sum of metabolites of di(2-ethylhexyl) phthalate, *ICC* intraclass correlation coefficient

[a]Adjusted for gestational age, weight gain, time since awoke, time slept, time since last food/drink, and time since last void

[b]When ICCs for ∑DEHP metabolites were not available, the mean ICC for the four individual metabolites was reported

## Potential sources of variability in urinary concentrations of phthalate metabolites and phenols

Unadjusted urinary concentrations of phthalate metabolites and phenols at each time point, the number of samples with concentrations below the limit of detection for the target biomarkers, and the number of missing concentrations are presented in Table 2. Specific gravity adjusted urinary concentrations for the same biomarkers and time points are available in Supplementary Table S1. Benzophenone-3 and triclosan were not included in further analyses because of the number of missing concentrations, 14% and 10%, respectively. The univariate relationship between the time-varying covariates and the phenols and phthalate metabolites are presented in Table 3. Associations were inconsistent across biomarkers and covariates. For example, later gestation was associated with lower 2,4-dichlorophenol concentrations (−2.5%, 95% CI: −4.4, −0.7) but higher 2,5-dichlorophenol concentrations (3.0%, 95% CI: 0.7, 5.4). Greater gestational weight gain was associated with higher BPA concentrations (2.5%, 95% CI: 0.4, 4.5) but did not have a significant impact on other study biomarkers. Although not all associations were significant, time slept prior to urine collection was generally associated with higher phenol but lower phthalate metabolite concentrations. Time since last void prior to urine collection was typically associated with higher biomarker concentrations, but again, not all associations were statistically significant.

## Reproducibility of urinary phenol concentrations

In our study population, urinary BPA concentrations were the least reproducible across 6 pregnancy samples (ICC: 0.24, 95% CI: 0.10, 0.49), while those of B-paraben and P-paraben were the most reproducible (ICC: 0.56, 95% CI: 0.36, 0.74 and ICC: 0.56, 95% CI: 0.36, 0.74, respectively). Adjustment for potential sources of variability in urinary concentrations (unrelated to exposure pathways) did not improve reproducibility (Table 4). The largest, yet modest, improvement was observed for 2,5-dichlorophenol; after adjustment, the ICC increased from 0.34 to 0.39. Similarly, for 2,5-dichlorophenol as well as 2,4-dichlorophenol concentrations, the sensitivity, specificity, and percentage of correctly classified adjusted values typically reflected potential improvement relative to unadjusted measures (Table 5). For the other phenols, adjustment had an inconsistent impact on the percentage of correct classification, sensitivity, and specificity (Table 5). For example, adjustment improved the percentage of correctly classified for BPA in the 32–33 week urine samples (respective increases from 84% to a range of adjusted values from 89% to 100%) but ranged from poorer to similar classification accuracy for the 16–18 week urine (the unadjusted value was 74% and the adjusted values were between 58% and 74%).

## Reproducibility of urinary concentrations of phthalate metabolites

Urinary MEP concentrations were the most reproducible over pregnancy (ICC: 0.68; 95% CI: 0.50, 0.82) and reproducibility was essentially unchanged by adjustment (ICC: 0. 66; 95% CI: 0. 46, 0. 81) (Table 4). Adjustment did not impact the sensitivity of urinary MEP concentrations, which averaged 0.81 (range of unadjusted sensitivity values: 0.67–1.00) (Table 5).

In contrast to MEP, concentrations of ∑DEHP metabolites were not as reproducible (ICC: 0.32; 95% CI: 0.15, 0.55), and adjustment for potential sources of variation did not improve the ICC (Table 4). Additionally, the sensitivity, specificity, and percentage of correct classification for

**Table 5** Sensitivity, specificity, and percent correctly classified (as high vs. low average exposure across pregnancy[a]) using biomarker concentrations in a single pregnancy urine

| | 16–18 weeks | | 20–22 weeks | | 23–26 weeks | | 27–30 weeks | | 32–33 weeks | | 35–36 weeks | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Unadjusted | Adjusted[b] | Unadjusted | Adjusted[b] | Unadjusted | Adjusted[b] | Unadjusted | Adjusted[b] | Unadjusted | Adjusted[b] | Unadjusted | Adjusted[b] |
| Bisphenol A (μg/L) | | | | | | | | | | | | |
| % Crct | 74% | 58–74% | 74% | 63–74% | 74% | 58–79% | 74% | 79–89% | 84% | 89–100% | 74% | 58–74% |
| Sens/Spec | 0.57/0.83 | 0.33–0.57/0.69–0.83 | 0.57/0.83 | 0.43–0.57/0.75–0.83 | 0.57/0.83 | 0.33–0.67/0.69–0.85 | 0.57/0.83 | 0.67–0.83/0.85-0.92 | 0.71/0.92 | 0.83–1.0/0.92–1.0 | 0.57/0.83 | 0.33–0.57/0.69–0.83 |
| 2 4-Dichlorophenol (μg/L) | | | | | | | | | | | | |
| % Crct | 84% | 84–89% | 84% | 63–79% | 84% | 79–100% | 84% | 89–95% | 63% | 68–84% | 74% | 63–84% |
| Sens/Spec | 0.71/0.92 | 0.71–0.83/0.92–0.92 | 0.71/0.92 | 0.43–0.67/0.75–0.85 | 0.71/0.92 | 0.67–1.0/0.85–1.0 | 0.71/0.92 | 0.83–0.86/0.92–1.0 | 0.43/0.75 | 0.50–0.71/0.77–0.92 | 0.57/0.83 | 0.43–0.71/0.75–0.92 |
| 2 5-Dichlorophenol (μg/L) | | | | | | | | | | | | |
| % Crct | 84% | 68–89% | 84% | 89–100% | 95% | 79–95% | 84% | 74–89% | 63% | 58–79% | 74% | 68–79% |
| Sens/Spec | 0.71/0.92 | 0.50–0.83/0.77–0.92 | 0.71/0.92 | 0.83–1.0/0.92–1.0 | 0.86/1.00 | 0.67–0.86/0.85–1.0 | 0.71/0.92 | 0.57–0.83/0.83–0.92 | 0.43/0.75 | 0.33–0.67/0.69–0.85 | 0.57/0.83 | 0.50–0.67/0.77–0.85 |
| Butyl paraben (μg/L) | | | | | | | | | | | | |
| % Crct | 74% | 58–74% | 84% | 63–89% | 74% | 63–89% | 84% | 68–84% | 84% | 79–100% | 84% | 79–100% |
| Sens/Spec | 0.57/0.83 | 0.33–0.57/0.69–0.83 | 0.71/0.92 | 0.43–0.83/0.75–0.92 | 0.57/0.83 | 0.43–0.83/0.75–0.92 | 0.71/0.92 | 0.50-0.71/0.77–0.92 | 0.71/0.92 | 0.67–1.0/0.85–1.0 | 0.71/0.92 | 0.67–1.0/0.85–1.0 |
| Methyl paraben (μg/L) | | | | | | | | | | | | |
| % Crct | 83% | 61–83% | 79% | 74–89% | 79% | 68–84% | 79% | 74–89% | 56% | 44–56% | 89% | 68–84% |
| Sens/Spec | 0.67/0.92 | 0.33–0.67/0.75–0.92 | 0.67/0.85 | 0.57–0.83/0.83–0.92 | 0.67/0.85 | 0.50–0.71/0.77–0.92 | 0.67/0.85 | 0.57–0.83/0.83–0.92 | 0.20/0.69 | 0.00–0.29/0.62–0.73 | 0.83/0.92 | 0.50–0.71/0.77–0.92 |
| Propyl paraben (μg/L) | | | | | | | | | | | | |
| % Crct | 84% | 63–79% | 84% | 79–95% | 74% | 68–89% | 74% | 74–95% | 74% | 58–79% | 74% | 68–89% |
| Sens/Spec | 0.71/0.92 | 0.43–0.67/0.75–0.85 | 0.71/0.92 | 0.67–0.86/0.85–1.0 | 0.57/0.83 | 0.50–0.83/0.77–0.92 | 0.57/0.83 | 0.57–0.86/0.83–1.0 | 0.57/0.83 | 0.33–0.67/0.69–0.85 | 0.57/0.83 | 0.50–0.83/0.77–0.92 |
| Monoethyl phthalate (μg/L) | | | | | | | | | | | | |
| % Crct | 79% | 68–79% | 79% | 74–79% | 100% | 89–100% | 89% | 84–95% | 89% | 74–95% | 89% | 79–95% |
| Sens/Spec | 0.67/0.85 | 0.50–0.67/0.77–0.85 | 0.67/0.85 | 0.57–0.67/0.83–0.85 | 1.00/1.00 | 0.83–1.0/0.92–1.0 | 0.83/0.92 | 0.71–0.86/0.92–1.0 | 0.83/0.92 | 0.57–0.86/0.83–1.0 | 0.83/0.92 | 0.67–0.86/0.85–1.0 |
| ∑DEHP metabolites (μmol/L) | | | | | | | | | | | | |
| % Crct | 79% | 63–79% | 79% | 68–74% | 89% | 79–89% | 79% | 58–74% | 79% | 68–74% | 79% | 63–89% |
| Sens/Spec | 0.67/0.85 | 0.43–0.67/0.75–0.85 | 0.67/0.85 | 0.50–0.57/0.77–0.83 | 0.83/0.92 | 0.67–0.83/0.85–0.92 | 0.67/0.85 | 0.33–0.57/0.69–0.83 | 0.67/0.85 | 0.50–0.57/0.77–0.83 | 0.67/0.85 | 0.43–0.83/0.75–0.92 |

*∑DEHP metabolites* sum of metabolites of di(2-ethylhexyl) phthalate, *% Crct* percentage of correctly classified, *Sens* sensitivity, *Spec* specificity

[a]Estimated using the mean of the residuals for the six urine samples collected across pregnancy

[b]Adjusted analyses use imputed variables and results presented are the range of values across the ten imputed datasets. Estimates are adjusted for gestational age, weight gain, time since awoke, time slept, time since last food/drink, and time since last void

∑DEHP concentrations were not improved by adjustment and, instead, often were worse (Table 5).

## Discussion

Developing cost-effective approaches for accurately characterizing exposure to phthalates or phenols during prenatal development is important for conducting epidemiologic studies. Our goal was to improve reproducibility of these measures over time by minimizing potential sources of time-varying within-person variation in these biomarkers unrelated to exposure. We hoped to thereby enhance their utility in studies of human health impacts as well as in studies identifying key exposure risk factors. We hypothesized that we could achieve this goal by implementing an analytic strategy to decrease random temporal variability. To the best of our knowledge, this goal has not been addressed in previous studies that have characterized variability in urine EDC measures [11, 21] but have not attempted to minimize that variability. The one exception is an analysis adjusting for time-varying sampling conditions (e.g., hour of random urine collection) as well as urine handling (e.g., storage time prior to freezing), but reproducibility was not improved by this approach and the analysis did not consider the role of individual time-varying behaviors [22].

In addition, our study is notable for having first morning urine samples available for 6 time points across pregnancy, whereas previous studies in pregnancy have typically collected urine at no more than 2–4 time points [9–11]. Our design optimized the ability to characterize long-term pregnancy exposure in ways not done previously. In addition, because the study population's sociodemographic characteristics were relatively homogeneous, potential variability related to socio-demographics was minimized. Lastly, first morning voids were used to optimize the reproducibility of our measures as timed collections are more likely to be comparable across participants vis-a-vis proximity to meal time, personal grooming activities, and last void than random urine collections. For example, most, if not all, previous studies of the reproducibility of MEP in pregnant women have not used first morning urine samples [11, 22–26]. Our relatively strong ICCs for phthalates (e.g., MEP and ∑DEHP metabolites) compared to other studies may, in part, reflect the use of a first morning timed urine collection (Table 4). Despite these design strengths, adjustment for *a priori* potential sources of variability unrelated to exposure pathways had minimal impact on the reproducibility of urinary concentrations of phthalate metabolites or phenols during pregnancy (Table 4). Similarly, the specificity, sensitivity, and probability of correctly classifying urinary concentrations across pregnancy using a single sample were not materially improved by adjustment for these potential sources of variability. This pattern of findings applied to the biomarkers of both phthalates (MEP, ∑DEHP metabolites) and phenols ($n = 6$) we assessed (Table 5). One possible explanation for our findings is that identifying correlates of urinary concentrations of these compounds may be particularly difficult in pregnant women for whom changes in xenobiotic metabolism, body composition, nutritional status, and even health behaviors may contribute to enhanced variability in exposure to and absorption, distribution, metabolism, and elimination of phenols and phthalates [11, 27]. Although use of first morning voids (rather than random collections) may have improved reproducibility by decreasing variability in urine collection circumstances, a timed collection also could limit the variability of key covariates in our analysis (e.g., time since last void) and thus minimize the impact of our covariate adjustments. However, despite collection of first morning voids, there was variability in measures related to urine sampling time. For example, across the 114 urine collections, time since last void ranged from 1.3 to 12.8 h with a mean (SD) of 6.2 (2.5) hours and time since last food/drink ranged from 10 min to 15.0 h with a mean (SD) of 8.3 (3.2) hours (data not shown), consistent with a pregnant population where awakenings to void or eat (followed by return to sleep) are not uncommon.

More fundamentally, our inability to improve the value of a urinary concentration from a single sample for predicting long-term exposure to the target biomarkers may reflect, in part, the uncertainty in identifying sources of variation unrelated to exposure regardless of pregnancy status. In addition, by design, we did not adjust for personal behaviors related to likely exposure pathways (e.g., personal care product use, diet). These behaviors may account for both within- and between-person variation in biomarker levels over time and thus could explain persistence of limited reproducibility of individual urine measures. Instead, our choice of adjustment factors was based on *a priori* considerations unrelated to exposure including the short half-life of the target biomarkers (e.g., time since last void would thereby impact urine concentrations) and the likely correlation of urinary concentrations with daily activities. For example, meal time may represent a period of potentially high exposure but is not specific enough to capture actual exposure. Despite these considerations, none of the *a priori* factors we considered consistently improved the predictive utility of a single urinary concentration of any of the target biomarkers. This result is consistent with the relatively modest associations observed between the urine EDC biomarker concentrations and many of our time-varying covariates (Table 3).

It is possible that sources of variation (unrelated to exposure pathways) may differ by exposure and population

characteristics. However, our findings are in a population with urinary concentrations of phthalate metabolites and phenols generally comparable to other population-based studies of pregnant women (Table 2) [9, 10, 23, 28–43]. For example, our average urinary concentrations across pregnancy for BPA, 2,4-dichlorophenol, M-paraben, P-paraben, triclosan, ∑DEHP metabolites, and MEP were within range of previously published studies of pregnant women (Table 2). Our urinary concentrations of 2,5-dichlorophenol were slightly lower, whereas benzophenone-3 concentrations were slightly higher (Table 2), suggesting that use of products with sunscreen containing benzophenone-3 may be more common in our largely non-Hispanic white population than observed in other studies. More generally, our findings are consistent with lower 2,5-dichlorophenol and higher benzophenone-3 concentrations observed in white non-Hispanic populations as compared to other racial or ethnic groups [44, 45]. Although we were unable to improve the value of predicting longer-term exposure from a single sample for urinary concentration of phthalate metabolites or phenols, we demonstrated good reproducibility of urinary MEP concentrations (ICC = 0.68) and reasonable reproducibility of paraben concentrations (ICC = 0.44–0.56). In fact, concentrations of MEP as well as ∑DEHP metabolites typically had better reproducibility (unadjusted ICCs of 0.68 and 0.32, respectively) than has been observed in other studies (ICCs ranging from 0.21 to 0.50 and from 0.08 to 0.31, respectively) [11, 23–26]. For other biomarkers, reproducibility, whether adjusted or not, was similar to values reported elsewhere [4, 9–12, 25, 46–48]. For example, the ICC for M-paraben in our study was 0.44 and in other populations of pregnant women the ICC for M-paraben was between 0.24 and 0.61 (Table 4) [9, 10, 12, 48]. In contrast, our urinary 2,5-dichlorophenol concentrations had poorer reproducibility than observed elsewhere (unadjusted ICC of 0.34 vs. 0.49–0.61) (Table 4) [9, 10].

There are few studies assessing the sensitivity or specificity of biomarker concentrations from a single urine specimen for classifying high vs. low exposure across pregnancy (a.k.a., "surrogate analyses"), and to the best of our knowledge, there are no published reports in pregnant or non-pregnant adults assessing 2,4-dichlorophenol or 2,5-dichlorophenol. In our study population, the unadjusted specificity and percentage of correctly classified for these two phenols were relatively high (for 2,4-dichlorophenol: 0.75–0.92 and 63–84% and for 2,5-dichlorophenol: 0.75–1.0 and 63–95%, respectively) thereby providing a point of comparison for future studies in other populations. In contrast to other phenols examined, likely exposure sources for 2,4-dichlorophenol and 2,5-dichlorophenol or their precursors are not predominantly food or personal care products but rather from herbicides (2,4-dichlorophenol), by-products of chlorinated chemical manufacture (2,4-dichlorophenol), or consumer goods such as mothballs and toilet bowl deodorants (2,5-dichlorophenol) [44]. These specific exposure risk factors may impact the reproducibility and predictive value of urine 2,4-dichlorophenol and 2,5-dichlorophenol as biomarkers across populations.

To the best of our knowledge, there are only four published reports of surrogate analyses in pregnant women on the other biomarkers we studied. These include women attending a fertility clinic in Massachusetts [11, 48], a pregnancy cohort in New York City [23], and a study of personal-care product use in pregnant women in Ottawa, Canada [34]. Depending on the study, high vs. low exposures for surrogate analyses have been variously defined either based on quantiles of the observed urinary concentrations or on reference population data (e.g., National Health and Nutrition Examination Survey) [23]. For BPA, M-paraben, P-paraben, B-paraben, MEP, and individual DEHP metabolites or their sum, previously reported sensitivities and specificities among pregnant women are similar to ours [11, 23, 34, 48]. For example, in our analysis the (unadjusted) sensitivity and specificity for BPA were between 0.57 and 0.71 and between 0.83 and 0.92, respectively, which is comparable to other populations of pregnant women (sensitivity range: 0.60–0.70; specificity range: 0.66–0.85) [11, 34]. Similarly, in our study, the range of sensitivity values for MEP (0.67–1.0) and P-paraben (0.57–0.71) were similar to previous studies (MEP: 0.62–0.81; P-paraben: 0.63–0.73), while specificity values (MEP: 0.85–1.0; P-paraben: 0.83–0.92) were on the higher end of previous reports (MEP 0.43–0.90; P-paraben: 0.80–0.86) [11, 23, 34].

In our surrogate analysis, we used the mean of six urinary concentrations as the "gold standard" measure, assuming it represented exposure over the entire pregnancy; however, it is possible that there was variability in exposures over pregnancy that may not have been captured even with six samples. Additionally, when assessing how well a single urinary concentration predicted overall pregnancy concentrations (using the six samples), the single urinary concentration was included in the mean concentration of the six samples. We chose this approach to optimize our ability to characterize exposure across pregnancy. However, this approach also means the single urinary concentration is not completely independent of the "gold standard." Moreover, with a modest sample size of 19, only 1 woman with discordant high/low exposure ranking based on the single compared to the mean of six urine samples, could have a substantial impact on our calculated sensitivity and specificity. For example, seven women had BPA concentrations in the highest tertile when using the mean of six samples (the "gold standard"); among these seven women, four spot urinary concentrations at 16–18 weeks were correctly classified as "high" and three were incorrectly classified as

"low/medium", resulting in a sensitivity of 57%. If one woman moved from the "high" to the "low/medium" group, the sensitivity would be reduced to 43%. Thus our findings are limited by sample size and the resultant sensitivity to changes in ranking for individual observations. That said, our findings are remarkably consistent with the existing literature, and in general, the range of possible values did not alter our conclusion that adjustment for variability in individual behavior over time did not substantially improve sensitivity or specificity.

In summary, we optimized our ability to assess long-term exposure over pregnancy with multiple ($n = 6$) timed urine collections in a relatively homogeneous population of women. Collection of detailed diaries at each time period allowed us to assess and adjust for potential sources of variability related to time-varying behaviors (e.g., hours slept, last void). Despite these design strengths, adjustment for variability (unrelated to exposure) had minimal impact on the reproducibility, sensitivity, specificity, or percentage of correct classification of urinary concentrations of phenols or phthalate metabolites during pregnancy. Thus, despite using a demographically homogeneous study population, this approach does not appear to enhance the utility of a single urine measure for assessing exposure in pregnancy. Certain urinary biomarkers had higher ICCs (e.g., MEP), suggesting that they may be more reliably measured with just one sample than other biomarkers. Future studies need to carefully consider the exposure of interest and whether it is appropriate to use biomarker concentrations from a single spot urine sample to represent exposures over pregnancy given the exposure timeframe and the reproducibility of the biomarker.

## Disclaimer

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.

**Acknowledgements** We thank the study participants for their time and participation which made this work possible. We also thank Manori Silva, Prabha Dwivedi, Ella Samandar, Jim Preau, Xiaoliu Zhou, and Tao Jia for technical assistance in measuring the urinary concentrations of phenols and phthalate metabolites.

**Funding** Support for this research was provided by the following grants: RD-83459301 and RD-835434010 from the U.S. Environmental Protection Agency; P20 ES018163, P20 ES018163-S1, P01 ES022848, and R01 ES014864 from the National Institute of Environmental Health Sciences (NIEHS/NIH). MMY was supported by NIEHS training grant T32 ES007069.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Centers for Disease Control and Prevention. Fourth report on human exposure to environmental chemicals, updated tables. US Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA. Retrieved from: https://www.cdc.gov/biomonitoring/pdf/FourthReport_UpdatedTables_Feb2015.pdf (accessed on May 20, 2015).
2. Diamanti-Kandarakis E, Bourguignon J-P, Giudice LC, Hauser R, Prins GS, Soto AM, et al. Endocrine-disrupting chemicals: an Endocrine Society Scientific Statement. Endocr Rev. 2009;30:293–342.
3. Meeker JD. Exposure to environmental endocrine disruptors and child development. Arch Pediatr Adolesc Med. 2012;166:952–8.
4. Braun JM, Kalkbrenner AE, Calafat AM, Bernert JT, Ye X, Silva MJ, et al. Variability and predictors of urinary bisphenol A concentrations during pregnancy. Environ Health Perspect. 2011;119:131.
5. Centers for Disease Control and Prevention. Biomonitoring Summary: Parabens. US Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA. Retrieved from: http://www.cdc.gov/biomonitoring/Parabens_BiomonitoringSummary.html (accessed on October 5, 2015).
6. Centers for Disease Control and Prevention. Biomonitoring Summary: 2,4-Dichlorophenol. US Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA. Retrieved from: http://www.cdc.gov/biomonitoring/24D_BiomonitoringSummary.html (accessed on October 5, 2015).
7. Centers for Disease Control and Prevention. Biomonitoring Summary: 2,5-Dichlorophenol. US Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA. Retrieved from: http://www.cdc.gov/biomonitoring/25D_BiomonitoringSummary.html (accessed on October 5, 2015).
8. Wang L, Asimakopoulos AG, Kannan K. Accumulation of 19 environmental phenolic and xenobiotic heterocyclic aromatic compounds in human adipose tissue. Environ Int. 2015;78:45–50.
9. Kim S, Choi K. Occurrences, toxicities, and ecological risks of benzophenone-3, a common component of organic sunscreen products: a mini-review. Environ Int. 2014;70:143–57.
10. Philippat C, Wolff MS, Calafat AM, Ye X, Bausell R, Meadows M, et al. Prenatal exposure to environmental phenols: concentrations in amniotic fluid and variability in urinary concentrations during pregnancy. Environ Health Perspect. 2013;121:1225.
11. Meeker JD, Cantonwine DE, Rivera-González LO, Ferguson KK, Mukherjee B, Calafat AM, et al. Distribution, variability, and predictors of urinary concentrations of phenols and parabens among pregnant women in Puerto Rico. Environ Sci Technol. 2013;47:3439–47.
12. Braun JM, Smith KW, Williams PL, Calafat AM, Berry K, Ehrlich S, et al. Variability of urinary phthalate metabolite and bisphenol A concentrations before and during pregnancy. Environ Health Perspect. 2012;120:739.
13. Guidry VT, Longnecker MP, Aase H, Eggesbø M, Zeiner P, Reichborn-Kjennerud T, et al. Measurement of total and free urinary phenol and paraben concentrations over the course of pregnancy: assessing reliability and contamination of specimens

in the Norwegian mother and child cohort study. Environ Health Perspect. 2015;123:705–11.
14. Calafat AM, Ye X, Wong L-Y, Reidy JA, Needham LL. Exposure of the US population to bisphenol A and 4-tertiary-octylphenol: 2003-4. Environ Health Perspect. 2008;116:39–44.
15. Kato K, Silva MJ, Needham LL, Calafat AM. Determination of 16 phthalate metabolites in urine using automated sample preparation and on-line preconcentration/high-performance liquid chromatography/tandem mass spectrometry. Anal Chem. 2005;77: 2985–91.
16. Silva MJ, Samandar E, Preau JL, Reidy JA, Needham LL, Calafat AM. Quantification of 22 phthalate metabolites in human urine. J Chromatogr B. 2007;860:106–12.
17. Ye X, Kuklenyik Z, Needham LL, Calafat AM. Automated on-line column-switching HPLC-MS/MS method with peak focusing for the determination of nine environmental phenols in urine. Anal Chem. 2005;77:5407–13.
18. Ye X, Kuklenyik Z, Bishop AM, Needham LL, Calafat AM. Quantification of the urinary concentrations of parabens in humans by on-line solid phase extraction-high performance liquid chromatography–isotope dilution tandem mass spectrometry. J Chromatogr B. 2006;844:53–9.
19. Duty SM, Ackerman RM, Calafat AM, Hauser R. Personal care product use predicts urinary concentrations of some phthalate monoesters. Environ Health Perspect. 2005;113:1530–5.
20. SAS Institute Inc. Release version 9.1.3. SAS Institute Inc, Cary, NC. 2000–2004.
21. Shoukri MM, Donner A, El-Dali A. Covariate-adjusted confidence interval for the intraclass correlation coefficient. Contemp Clin Trials. 2013;36:244–53.
22. Cox KJ, Porucznik CA, Anderson DJ, Brozek EM, Szczotka KM, Bailey NM, et al. Exposure classification and temporal variability in urinary bisphenol A concentrations among couples in Utah—the HOPE study. Environ Health Perspect. 2016; 124:498.
23. Mortamais M, Chevrier C, Philippat C, Petit C, Calafat AM, Ye X, et al. Correcting for the influence of sampling conditions on biomarkers of exposure to phenols and phthalates: a 2-step standardization method based on regression residuals. Environ Health. 2012;11:29.
24. Adibi JJ, Whyatt RM, Calafat AM, Camann D, Nelson H, Bhat HK, et al. Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. Environ Health Perspect. 2008;116:467–73.
25. Irvin EA, Calafat AM, Silva MJ, Aguilar-Villalobos M, Needham LL, Hall DB, et al. An estimate of phthalate exposure among pregnant women living in Trujillo, Peru. Chemosphere. 2010;80: 1301–7.
26. Casas M, Valvi D, Ballesteros-Gomez A, Gascon M, Fernández MF, Garcia-Esteban R, et al. Exposure to bisphenol A and phthalates during pregnancy and ultrasound measures of fetal growth in the INMA-Sabadell cohort. Environ Health Perspect. 2016;124:521–8.
27. Ferguson KK, McElrath TF, Ko Y-A, Mukherjee B, Meeker JD. Variability in urinary phthalate metabolite levels across pregnancy and sensitive windows of exposure for the risk of preterm birth. Environ Int. 2014;70:118–24.
28. Abduljalil K, Furness P, Johnson TN, Rostami-Hodjegan A, Soltani H. Anatomical, physiological and metabolic changes with gestational age during normal pregnancy. Clin Pharmacokinet. 2012;51:365–96.
29. Casas L, Fernández MF, Llop S, Guxens M, Ballester F, Olea N, et al. Urinary concentrations of phthalates and phenols in a population of Spanish pregnant women and children. Environ Int. 2011;37:858–66.
30. Castorina R, Bradman A, Fenster L, Barr DB, Bravo R, Vedar MG, et al. Comparison of current-use pesticide and other toxicant urinary metabolite levels among pregnant women in the CHAMACOS cohort and NHANES. Environ Health Perspect. 2010;118:856.
31. Mortensen ME, Calafat AM, Ye X, Wong L-Y, Wright DJ, Pirkle JL, et al. Urinary concentrations of environmental phenols in pregnant women in a pilot study of the National Children's Study. Environ Res. 2014;129:32–38.
32. Philippat C, Mortamais M, Chevrier C, Petit C, Calafat AM, Ye X, et al. Exposure to phthalates and phenols during pregnancy and offspring size at birth. Environ Health Perspect. 2012;120: 464–70.
33. Wolff MS, Engel SM, Berkowitz GS, Ye X, Silva MJ, Zhu C, et al. Prenatal phenol and phthalate exposures and birth outcomes. Environ Health Perspect. 2008;116:1092.
34. Bertelsen RJ, Engel SM, Jusko TA, Calafat AM, Hoppin JA, London SJ, et al. Reliability of triclosan measures in repeated urine samples from Norwegian pregnant women. J Expo Sci Environ Epidemiol. 2014;24:517–21.
35. Fisher M, Arbuckle TE, Mallick R, LeBlanc A, Hauser R, Feeley M, et al. Bisphenol A and phthalate metabolite urinary concentrations: daily and across pregnancy variability. J Expo Sci Environ Epidemiol. 2015;25:231–9.
36. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003-4. Environ Health Perspect. 2011;119:878.
37. Berman T, Hochner-Celnikier D, Calafat AM, Needham LL, Amitai Y, Wormser U, et al. Phthalate exposure among pregnant women in Jerusalem, Israel: results of a pilot study. Environ Int. 2009;35:353–7.
38. Bornehag C-G, Carlstedt F, Jönsson BA, Lindh C, Jensen TK, Bodin A, et al. Prenatal phthalate exposures and anogenital distance in Swedish boys. Environ Health Perspect. 2015;123: 101–7.
39. de Renzy-Martin KT, Frederiksen H, Christensen JS, Kyhl HB, Andersson A-M, Husby S, et al. Current exposure of 200 pregnant Danish women to phthalates, parabens and phenols. Reproduction. 2014;147:443–53.
40. Swan SH, Sathyanarayana S, Barrett ES, Janssen S, Liu F, Nguyen RHN, et al. First trimester phthalate exposure and anogenital distance in newborns. Human Reprod. 2015;30: 963–72.
41. Swan SH, Main KM, Liu F, Stewart SL, Kruse RL, Calafat AM, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ Health Perspect. 2005;113: 1056–61.
42. Ye X, Pierik FH, Hauser R, Duty S, Angerer J, Park MM, et al. Urinary metabolite concentrations of organophosphorous pesticides, bisphenol A, and phthalates among pregnant women in Rotterdam, the Netherlands: the Generation R study. Environ Res. 2008;108:260–7.
43. Yolton K, Xu Y, Strauss D, Altaye M, Calafat AM, Khoury J. Prenatal exposure to bisphenol A and phthalates and infant neurobehavior. Neurotoxicology Teratol. 2011;33: 558–66.
44. Arbuckle TE, Marro L, Davis K, Fisher M, Ayotte P, Bélanger P, et al. Exposure to free and conjugated forms of bisphenol A and triclosan among pregnant women in the MIREC cohort. Environ Health Perspect. 2015;123:277.
45. Ye X, Lee-Yang W, Zhou X, Calafat AM. Urinary concentrations of 2, 4-dichlorophenol and 2, 5-dichlorophenol in the US population (National Health and Nutrition Examination Survey, 2003-10): trends and predictors. Environ Health Perspect. 2014; 122:351.

46. Calafat AM, Wong L-Y, Ye X, Reidy JA, Needham LL. Concentrations of the sunscreen agent benzophenone-3 in residents of the United States: National Health and Nutrition Examination Survey 2003-4. Environ Health Perspect. 2008;116:893.
47. Jusko TA, Shaw PA, Snijder CA, Pierik FH, Koch HM, Hauser R, et al. Reproducibility of urinary bisphenol A concentrations measured during pregnancy in the Generation R Study. J Expo Sci Environ Epidemiol. 2014;24:532–6.
48. Cantonwine DE, Ferguson KK, Mukherjee B, McElrath TF, Meeker JD. Urinary bisphenol A levels during pregnancy and risk of preterm birth. Environ Health Perspect. 2015;123:895–901.
49. Smith KW, Braun JM, Williams PL, Ehrlich S, Correia KF, Calafat AM, et al. Predictors and variability of urinary paraben concentrations in men and women, including before and during pregnancy. Environ Health Perspect. 2012;120:1538–43.