Variability of Organophosphorous Pesticide Metabolite Levels in Spot and 24-hr Urine Samples Collected from Young Children during 1 Week

Asa Bradman, University of California Berkeley
Katherine Kogut, University of California Berkeley
Ellen A. Eisen, University of California Berkeley
Nicholas P. Jewell, University of California Berkeley
Lesliam Quiros-Alcala, University of California Berkeley
Rosemary Castorina, University of California Berkeley
Jonathan Chevrier, University of California Berkeley
Nina T. Holland, University of California Berkeley
Dana Barr, Emory University
Geri Kavanagh-Baird, University of California Berkeley

Only first 10 authors above; see publication for full author list.

Journal Title: Environmental Health Perspectives
Volume: Volume 121, Number 1
Publisher: National Institute of Environmental Health Sciences (NIEHS) | 2013-01-01, Pages 118-124
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1289/ehp.1104808
Permanent URL: https://pid.emory.edu/ark:/25593/s5c08

Final published version: http://dx.doi.org/10.1289/ehp.1104808

Copyright information:
Publication of EHP lies in the public domain and is therefore without copyright. All text from EHP may be reprinted freely.

Accessed November 12, 2017 10:03 PM EST
Spot urine samples have been used to assess organophosphorous (OP) pesticide exposures in epidemiological and biomonitoring studies, with samples analyzed for either nonspecific dialkyl phosphatase (DAP) metabolites or OP-specific metabolites (Fenske et al. 2002). Because collection of spot urine samples is relatively simple and noninvasive, this sample type is commonly used in studies involving children (Fenske et al. 2005). Several studies suggest that spot samples provide meaningful measures of exposure, reporting associations, for example, between residence near farmlands and OP metabolite levels in children’s urine (Fenske et al. 2000). However, concerns have been raised as to whether metabolites measured in a single spot sample accurately reflect exposure for the full day or beyond.

Several factors may result in exposure misclassification when using spot samples: a) Because the concentration of a chemical, urine volume, and rate of urinary excretion vary with fluid and salt intake, time of day, and other factors (Barr et al. 2005; Boeniger et al. 1993; Cornelis et al. 1996), exposure measurements may be more variable than other sample types, such as first morning voids (FMVs) (Kissel et al. 2005), which are more concentrated and reflect a longer period of accumulation, or 24-hr urine collections, considered by some as the “gold standard” (Aprea et al. 2002; Boeniger et al. 1993; Fenske and Elkner 1990; Martin et al. 1996); b) because of within-child variability in creatinine excretion (Barr et al. 2005; Boeniger et al. 1993), creatinine adjustment may introduce new variability that does not reflect urine dilution; and c) most OP pesticides have biological half-lives of only 12–36 hr (Needham 2005) and are metabolized and excreted from the body in hours to days (Barr 2008). Thus, OP pesticide metabolite excretion in a brief sampling period, whether measured in a spot, FMV, or 24-hr urine sample, may not accurately represent a child’s cumulative OP pesticide exposure over longer periods.

One recent study of 44 preschool-age children who provided between 10 and 26 biweekly spot urine samples over 21 months reported that within-child variability for the five DAP metabolites measured exceeded between-child variability by several factors (Griffith et al. 2011)—a finding that underscores concern about the reliability of this measure and raises questions about the source of within-child variability. To date, no studies have evaluated variability of OP metabolites in within-day samples or in 24-hr urine samples collected several days apart. Nor have studies compared variability in DAP measurements when expressed in terms of unadjusted concentrations, creatinine-adjusted concentrations, or urinary excretion rates.

To address these questions, in this study we aimed to a) evaluate the reproducibility of children’s nonspecific OP pesticide measurements within a 1-week period; b) evaluate the influence of concentration corrections (i.e., creatinine adjustment and use of urinary excretion rate) on reproducibility of DAP measurements; c) determine the degree of correspondence between pesticide metabolite concentration in spot urine samples (including FMVs) and same-day 24-hr samples; and d) evaluate the sensitivity and specificity of spot urine samples to classify full-week average measurements.

**Methods**

**Study population.** We recruited a convenience sample of 25 children (10 boys, 15 girls) from

Additional details about the study population, methods, and results are provided in the original publication.
clinics serving low-income families in the Salinas Valley, an agricultural region. Eligible children were between 3 and 6 years old, in good health with no history of diabetes or renal disease, toilet-trained, free of enuresis, and had English- or Spanish-speaking mothers who were at least 18 years old. Sampling occurred in March and April 2004. The study was approved by the University of California at Berkeley Committee for the Protection of Human Subjects, and parents provided written informed consent.

**Urine sample collection.** Each family participated for 7 consecutive days (study days 1–7). On day 1, study staff provided families with sampling supplies and instructions on sample collection. Supplies included specimen trays (cleaned, sterile Specian™; Baxter Scientific, McGaw Park, IL), gloves, collection jars with blank labels, a small refrigerator, and two 24-hr sampling record forms. On spot-sampling days (1, 3, 4, 6, and 7), families collected a single void at their convenience; families sometimes collected an FMV as the spot sample. On 24-hr sampling days (2 and 5), families were instructed to collect the child’s FMV, all daytime and evening spot voids, and the FMV of the following day as separate specimens, and to note the timing of all voids (including missed voids) during this period on the 24-hr sampling record form.

Sample collection methods have been previously validated in this age range (Lu et al. 2001). Children voided directly into a collection jar or into a clean, sterile specimen tray placed on a training potty or toilet. If samples were collected in specimen trays, parents transferred specimens into collection jars. Parents identified each sample as an FMV or a non-FMV spot sample, recorded the time of collection on jar labels, and stored specimens in the refrigerator until daily collection by research staff. On 24-hr sampling days, research staff reviewed the sampling record form with the families to assure its accuracy and completeness.

Twenty-four-hour samples included 3–11 collected voids (mean, 6.5). Overall, 86% (range, 50–100%) of the reported voids during the 24-hr sampling periods were collected. Twenty-two (44%) of the 24-hr samples were based on 100% collection of all voids. Reasons for missed voids included out-of-home bathroom use, toileting accidents, and participant error, such as missing an evening void. In sensitivity analyses, we excluded seven 24-hr samples with fewer than five total voids to confirm that collection errors did not substantially change study results (not shown).

**Sample processing and analysis.** Samples were processed at the Salinas Valley study office. The weight of each void was measured (grams), and the volume (milliliters) was estimated assuming a urine specific gravity of 1.022 (Salita et al. 1998). Individual voids from 24-hr sampling sessions that were not selected for individual analysis were pooled. After aliquotting, samples were stored at −80°C until being shipped on dry ice to the CDC for analysis.

We measured six DAP metabolites in all samples: three dimethyl (DM) phosphate metabolites—dimethylphosphate (DMP), dimethylthiophosphate (DMTP), and dimethyltrihalophosphate (DMDTP); and three diethyl (DE) phosphates—diethylphosphate (DEP), diethylthiophosphoramide (DETP), and diethyltrihalophosphite (DDETP). These metabolites represent approximately 81% of the agricultural OP pesticides used in the Salinas Valley (Bradman et al. 2007). Laboratory methods and quality control procedures are described in detail elsewhere (Bravo et al. 2004). Briefly, samples were lyophilized to remove water, redissolved in a 1:1 solution of acetonitrile and diethyl ether, and analyzed using gas chromatography–tandem mass spectrometry using isotope dilution. Creatinine concentrations were determined using a commercially available kit (Vitros CREA slides; Ortho Clinical Diagnostics, Raritan, NJ). Westgard rules for quality control were used to establish the validity of each analytical run (Westgard 2003). Limits of detection (LODs) were 0.2 μg/L for all DEs, 0.5 μg/L for DMP, 0.4 μg/L for DMTP, and 0.1 μg/L for DMDTP.

Levels below the LOD were imputed as LOD/2 (Hornung and Reed 1990), and molar concentrations were summed within each sample to yield total DM, total DE, and total DAP concentrations. Metabolite levels in 24-hr voids were computed as the volume-weighted average of concentrations in all samples collected on sampling days 2 and 5, which included the FMV samples from the following day. We expressed metabolite concentration and/or excretion in three ways: a) unadjusted metabolite concentration (nanomoles metabolite per liter of urine); b) creatinine-adjusted metabolite concentration (nanomoles metabolite per gram of creatinine); and c) urinary excretion rate (UER, in nanomoles metabolite per minute) computed as the total moles of excreted metabolite divided by the duration of time in minutes since the previous void. For 24-hr collections, time since the previous void was calculated as the time elapsed between the last void preceding the 24-hr sample collection (typically the last void on day 1 or 4) and the last void included in the 24-hr sample (typically the FMV on day 3 or 6).

**Data Analysis**

Statistical analyses were performed using Stata 10 for Windows (StataCorp LP, College Station, TX). We first computed descriptive statistics and assessed normality. Log₁₀-transformed values were used for subsequent analyses.

**Correspondence between spot samples and 24-hr samples.** We constructed generalized estimating equation (GEE) models using DAP concentration in a 24-hr sample as the outcome variable and concentration in a same-day spot sample (FMV or non-FMV) or a combination of these two spot samples as the predictor variable. Combinations of two spot samples were modeled as separate predictor variables and as a single variable computed as either the arithmetic or volume-weighted average of the individual samples. Analyses using each DAP variable type (i.e., unadjusted concentrations, creatinine-adjusted concentrations, and UER) were compared. Robust standard errors were calculated.

**Reproducibility of spot and 24-hr urine samples.** We used mixed random-effects models to compute the between-child and within-child variance for 24-hr samples from days 2 and 5, and for all FMV spot samples. For all non-FMV spot samples and for all spot samples (FMV and non-FMV), we estimated the variance attributable to each of three nested components—between-child, within-child, and within-child within-day variability (because multiple samples were available from the same day)—using two-level random-intercept models (Marchenko 2006). This method parses the total within-child variance into two parts: the component attributable to same-day within-child variability, and the component attributable to between-day-within-child variability (Hauser et al. 2004; Preau et al. 2010; Ye et al. 2011). We calculated Akaike Information Criterion (AIC) values for the nested models with the largest number of observations (the all spot samples model) to compare the fit of models using unadjusted DAP concentrations, creatinine-adjusted concentrations, and UER. We also calculated the intraclass correlation coefficient (ICC) for each group of samples. The ICC is the ratio of between-subject variance to total variance.

**Correlation of spot samples collected 0–6 days apart.** We computed Pearson’s correlations between pairs of samples collected on the same day (0 days apart) as well as on pairs of samples collected between 1 and 6 days apart. This analysis included all possible unique pairings of each child’s spot samples, and was conducted separately for pairs of FMV spots only, pairs of non-FMV spots only, and for all pairs of spots regardless of FMV status. Tests of significance were computed using a robust estimate of the variance.

**Sensitivity and specificity.** We evaluated the sensitivity and specificity of one, two, or three randomly-selected spot urine samples from each child as predictors of high (top 20%) or elevated (top 40%) weekly average DAP metabolite concentrations. For “true” exposure, we calculated the arithmetic mean metabolite concentration of all spot urine samples collected from each
child during the week, categorized these in quintiles, and assigned “true” high (top 20%) or elevated (top 40%) exposure level. For predictor sets, we created 10 data sets each containing one randomly selected spot sample per child (i.e., 25 observations total per data set). Within each randomly-selected data set, we again categorized the 25 metabolite concentration in quintiles, and assigned “predicted” high (top 20%) or elevated (top 40%) exposure levels. This process mimicked the exposure classification process in epidemiological studies, in which knowledge of the range and rankings of exposure in a population is often limited to what can be observed in the single sample or small number of samples collected per subject.

**Table 1.** Unadjusted and creatinine-adjusted DAP concentrations in urine samples.

<table>
<thead>
<tr>
<th>Type of sample</th>
<th>DF (%)</th>
<th>Unadjusted</th>
<th>Creatinine adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GM</td>
<td>Mean</td>
</tr>
<tr>
<td>Non-FMV spot samples (n = 137)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total DAPs</td>
<td>99.5</td>
<td>110</td>
<td>239</td>
</tr>
<tr>
<td>Total DMs</td>
<td>94.2</td>
<td>63.0</td>
<td>179</td>
</tr>
<tr>
<td>Total DEs</td>
<td>95.6</td>
<td>26.8</td>
<td>58.8</td>
</tr>
<tr>
<td>FMV samples (n = 110)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total DAPs</td>
<td>98.2</td>
<td>162</td>
<td>307</td>
</tr>
<tr>
<td>Total DMs</td>
<td>95.5</td>
<td>92.3</td>
<td>234</td>
</tr>
<tr>
<td>Total DEs</td>
<td>97.3</td>
<td>43.4</td>
<td>72.5</td>
</tr>
<tr>
<td>24-hr samples (n = 50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total DAPs</td>
<td>—</td>
<td>158.0</td>
<td>296</td>
</tr>
<tr>
<td>Total DMs</td>
<td>—</td>
<td>94.6</td>
<td>230</td>
</tr>
<tr>
<td>Total DEs</td>
<td>—</td>
<td>45.9</td>
<td>65.0</td>
</tr>
</tbody>
</table>

Abbreviations: DF, detection frequency; GM, geometric mean; Max, maximum.

The sensitivity and specificity figures we report represent the average sensitivity and specificity observed across the 10 separate random samples. We report findings separately for FMV samples only, for non-FMV samples only, and for all spot samples. To determine whether collection of more than one spot sample could improve sensitivity, we repeated this analysis using the arithmetic mean of two or three spot samples from each child collected on different days. Because some children lacked three FMV or non-FMV samples from different days, we only assessed the sensitivity and specificity of three spot samples in the “any spot sample” group.

**Results**

All children were Mexican American, and their ages ranged from 3 to 6.5 years (mean ± SD age, 4.5 ± 0.93 years). Children urinated between 3 and 12 times (mean, 5.7 voids) per 24-hr period, and the volume of their individual spot samples ranged from 4.8 to 642.2 mL (mean, 146.6 mL) for FMV samples and

![Figure 1](image-url)
from 8.4 to 238.3 mL (mean, 68.1 mL) for non-FMV spot samples. The creatinine concentration of spot samples ranged from 9.4 to 213.2 mg/dL (mean, 86.6 mg/dL) for FMV samples and from 4.5 to 158.7 mg/dL (mean, 68.6 mg/dL) for non-FMV samples. Creatinine concentration varied more within children than between children for both FMV (ICC = 0.37) and non-FMV (ICC = 0.22) spot samples. Individual DE and DM metabolites were detected in > 70% of samples, except for DE-DTP (17%) (data not shown). Total DM concentrations exceeded total DE concentrations, and total DAP levels reflected predominantly DM metabolites (Table 1).

Central tendency measures (e.g., means, medians) for unadjusted concentration values were similar for FMV spot samples and 24-hr samples, which were generally higher than levels in non-FMV spot samples. In contrast, central tendency measures for creatinine-adjusted concentration values were more similar for FMVs and non-FMV spot samples, with both generally lower than creatinine-adjusted 24-hr samples.

Figure 1 presents the total DAP metabolite concentrations [in log_{10} scale (nanomoles per gram creatinine)] for each participant for all spot samples collected over the 7 consecutive sampling days. There are no clear exposure trends over time among the children, though our data indicate that shifts of up to two orders of magnitude can occur over the week (Figure 1, P7) or within a single day (Figure 1, P19, day 5). Twenty-four-hour samples collected only 3 days apart can likewise differ by an order of magnitude (e.g., Figure 1, P2, P7, P25).

Table 2 presents results of GEE models examining how well same-day spot samples predict 24-hr metabolite levels. For models examining a single unadjusted spot sample (FMV or non-FMV) and its respective 24-hr sample, the predictive power of the model, defined by the coefficient of determination (R^2), was highest for FMVs [R^2 for total DAP metabolites (nmol/mL) = 0.54 for FMVs vs. 0.39 for non-FMV spots]. The predictive power of the model was higher for creatinine-adjusted metabolites compared to unadjusted metabolites, especially for non-FMVs, with little difference between the sample types (total DAP R^2 = 0.57 for FMV and 0.63 for non-FMVs). The predictive power was lowest for UER models.

The best-fitting models were obtained when the arithmetic mean of an FMV and a non-FMV sample was used to predict the 24-hr values (Table 2). The best model fit was observed for creatinine-adjusted values (total DAP R^2 = 0.84), but the variability in 24-hr metabolite levels explained by unadjusted values and UER estimates were comparable (R^2 = 0.79 and 0.77, respectively). Model fit was strongest for analyses of total DM metabolites, and weakest for total DE metabolites. Volume-weighted averages of FMV and non-FMV spots produced model fits that were similar to that obtained with the arithmetic mean (data not shown). Results of sensitivity analyses that excluded seven 24-hr samples with fewer than five spot samples were consistent with those described above (data not shown).

AIC values for mixed-random-effects models used to estimate the variance in total DAP attributable to between-child versus within-child sources based on all spot urine samples (n = 247) indicated that best model fit was achieved with creatinine-adjusted values (AIC = 364; Table 3) versus AIC = 372 and 451, respectively, for unadjusted concentration and UER models (see Supplemental Material, Table S1 [http://dx.doi.org/10.1289/ehp.1104808]). Relative to unadjusted concentration and UER models, use of creatinine-adjusted values also provided the greatest degree of distinction between individual children [e.g., ICC = 0.32 for creatinine-adjusted total DAP metabolites in the “any spot sample” model (Table 3) vs. ICC = 0.21 and 0.15, respectively, for the analogous unadjusted concentration and UER models (see Supplemental Material, Table S1)].
between a factor between five (e.g., 57% vs. 12% of the total variance for non-FMV spot samples) and eight (e.g., 88% vs. 12% of the total variance for FMV spot samples) for these metabolites.

On the basis of their superior model performance in the preceding analyses, we used creatinine-adjusted values in subsequent analyses. Table 4 presents Pearson correlations of creatinine-adjusted metabolite levels in spot samples collected 0–6 days apart. Correlations for total DAPs and total DMs were moderate (r = 0.5) and statistically significant for samples collected on the same day (i.e., 0 days apart) or 1 day apart, and became weaker as the number of days between samples increased. All correlations were weak and not significant for samples collected 5–6 days apart. DE metabolites had weaker correlations compared with DMs, and the decay in correlations across days was more rapid for DEs than for DM or total DAP metabolites. FMV samples tended to have higher correlations than non-FMV samples or any spot samples. The correlations between samples collected 2–3 days apart are comparable with Pearson correlations between levels in 24-hr samples collected 3 days apart, which were 0.35, 0.36, and 0.15, respectively, for total DAP, DM, and DE metabolites.

Table 5 presents results of sensitivity and specificity analyses. In this sample of participants, children with high (top 20%) weekly average total DAP metabolite concentrations would be correctly classified by a single non-FMV spot sample 52% of the time. Using two samples increased the sensitivity marginally to 56%. FMVs also showed marginally higher sensitivity than non-FMV samples. Within the “any spot sample group,” three-spot predictors offered little apparent advantage over two-spot predictors. Overall, the sensitivity was higher to classify children in the top 40th percentile (elevated) than in the top 20th percentile, with a slight improvement with two versus one sample. Specificity was uniformly higher than sensitivity. Use of a single non-FMV spot sample would correctly identify a child in the lower 80th percentile (i.e., not “high”) 88% of the time, and would identify a child in the lower 60th percentile (i.e. not “elevated”) 78% of the time.

**Discussion**

This study of variability in DAP metabolites of OP pesticides in urine samples from children 3–6 years old indicates high variability over a 1-week time frame. Although we observed strong correlations between full-day 24-hr samples and same-day spot samples (see Pearson correlations, Table 2), there was weak correlation between DAP metabolite levels in 24-hr samples collected 3 days apart or in spot samples collected > 1 day apart, with correlations weaker for DE compared with DM DAP metabolites. Further, within-child variance was approximately two to three times the between-child variance for total DAP metabolites (e.g., 66% vs. 34% for FMV samples for creatinine-adjusted total DAPs, ICC = 0.27–0.35). This ratio was lower (and thus, the ICC higher) in DM metabolites compared with DE metabolites (ICC = 0.30–0.39 vs. ICC = 0.11–0.16, respectively). Finally, we found that DAPs measured in one, two, or even three spot urine samples have relatively low sensitivity to identify children who would be considered the most highly exposed on the basis of their average full-week DAP concentrations. For example, a single non-FMV spot sample would correctly identify only 52% of the children whose true weekly exposure was in the top quintile, suggesting high type 2 classification errors; by contrast, spot samples appear to offer good specificity, with a single non-FMV spot sample correctly identifying about 88% of children with true total DAP exposure below the top quintile. Overall, the high within-child variance, the weak correlation across days, and low specificity suggests that single-day measurements may not adequately characterize exposure for longer time frames necessary for chronic risk assessments or epidemiologic studies.

Our analysis replicated methods applied by Ye et al. (2011) and Preau et al. (2010) to partition the variance components among between-child, within-child between-day, and within-child within-day components for sample types with more than one sample per day (Table 3). Similar to their findings for bisphenol A and monooethyl phthalate, we found that overall within-person variance was higher than between-person variance, and that within-day variance was the largest component of total variance, suggesting that differences in exposure between days were lower compared with fluctuations during the course of each day. This finding suggests that eliminating within-day variability would substantially reduce within-child relative to between-child variability overall. However, we found that 65% of the total variance in 24-hr urine samples collected 3 days apart was due to within-child between-day variability, even though the 24-hr samples reflect complete sampling with no within-day variability (i.e., the entire sample was collected). These apparently contradictory findings suggest to us that the variance component analyses may overestimate the within-person within-day variance component. Specifically, we suspect the models may overestimate the degree of within-day variability that is biologically plausible, given that the total number of possible within-day observations is capped by the number of times individuals urinate each day, which in this study ranged between 3 and 12 times.

Our study provides important information about the appropriateness of creatinine adjustment to control for urinary dilution in children 3–6 years of age. Overall, our findings suggest that creatinine adjustment of urinary

<table>
<thead>
<tr>
<th>Type of sample</th>
<th># Samples</th>
<th>Variance</th>
<th>Percent total variance</th>
<th>ICC</th>
<th>Variance</th>
<th>Percent total variance</th>
<th>ICC</th>
<th>Variance</th>
<th>Percent total variance</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-FMV spot samples</td>
<td>137</td>
<td>0.075</td>
<td>25</td>
<td>0.27b</td>
<td>0.114</td>
<td>27</td>
<td>0.30b</td>
<td>0.041</td>
<td>12</td>
<td>0.14b</td>
</tr>
<tr>
<td>Between child</td>
<td>0.105</td>
<td>34</td>
<td>0.149</td>
<td>35</td>
<td>0.198</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within child, between day</td>
<td>0.125</td>
<td>41</td>
<td>0.162</td>
<td>38</td>
<td>0.198</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within child, within day</td>
<td>0.105</td>
<td>34</td>
<td>0.172</td>
<td>38</td>
<td>0.39</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIV samples</td>
<td>110</td>
<td>0.207</td>
<td>66</td>
<td>0.276</td>
<td>62</td>
<td>0.246</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between child</td>
<td>0.105</td>
<td>34</td>
<td>0.276</td>
<td>62</td>
<td>0.246</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within child</td>
<td>0.125</td>
<td>41</td>
<td>0.162</td>
<td>38</td>
<td>0.198</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any spot samples (FMV or Non-FMV)</td>
<td>247</td>
<td>0.097</td>
<td>31</td>
<td>0.32b</td>
<td>0.157</td>
<td>36</td>
<td>0.37b</td>
<td>0.033</td>
<td>10</td>
<td>0.11b</td>
</tr>
<tr>
<td>Between child</td>
<td>0.064</td>
<td>20</td>
<td>0.070</td>
<td>16</td>
<td>0.049</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within child, between day</td>
<td>0.064</td>
<td>20</td>
<td>0.070</td>
<td>16</td>
<td>0.049</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within child, within day</td>
<td>0.152</td>
<td>49</td>
<td>0.213</td>
<td>48</td>
<td>0.238</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-hr voids</td>
<td>50</td>
<td>0.089</td>
<td>35</td>
<td>0.35</td>
<td>0.131</td>
<td>36</td>
<td>0.026</td>
<td>16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Within child</td>
<td>0.168</td>
<td>65</td>
<td>0.237</td>
<td>64</td>
<td>0.132</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aNo. of samples used in calculation. bRatio of between-child to total variability as calculated using a one-factor (child) as opposed to a two-factor nested mixed-effects model. cBecause FMV spots and 24-hr voids allow only one measure per day, the distinction between within-day versus between-day variability is not applicable.
DAP concentrations in children ages 3–6 years maximizes between-child relative to within-child variance (thus reducing exposure misclassification), improves estimation of the average 24-hr DAP concentration on the basis of spot urine samples, and may decrease the difference between FMV and easier-to-collect non-FMV spot samples in terms of their ability to estimate total 24-hr excretion. Despite this empirical evidence suggesting that creatinine adjustment effectively normalized the metabolite concentrations, this method may introduce other sources of variability. Specifically, creatinine varies as a function of sex, age, meat consumption, body size, and muscle mass (Boeniger et al. 1993). Thus, large differences in size, muscle mass, and diet between developing children of the same age could result in very different normalized metabolite values among children with the same exposure. In future analyses, we will examine creatinine excretion patterns in these children and assess adjustment for specific gravity (not yet quantified in these samples) as an alternative approach to account for urinary dilution.

In summary, our findings raise concerns about the utility of urinary DAP metabolites as a biomarker of chronic OP pesticide exposure in young children. Two recent published analyses from our Center have detected no association between DAP measures in children and adverse outcomes, although they did detect significant associations between maternal prenatal DAP measurements and child neurodevelopmental outcomes (Bouchard et al. 2011; Marks et al. 2010), despite the fact we have also observed high within versus between variability in maternal urine (Bradman et al. 2005). One interpretation of this finding is that OP exposures in childhood are less critical than exposures that occur in utero. The fetus is so exceptionally sensitive to OPs that the strength of the prenatal exposure effects are evident despite limitations in the exposure measure. The current study suggests another plausible explanation: that misclassification of child OP exposure biased results toward the null hypothesis.

Our study has several limitations. Collecting 24-hr urine samples from young children proved challenging, and missed voids due to occasional toileting accidents or other circumstances were beyond the control of study staff. Though 86% of the voids reported to have occurred during 24-hr sampling periods were collected, missed voids meant that individual spot samples made up a greater proportion of the 24-hr volume-weighted average than they should have. Thus, our estimates of the association between metabolites in spot and 24-hr samples (Table 2) represent an upper bound. We measured class-specific organophosphate pesticide DAP metabolites, and our findings should not be generalized to pesticide-specific OP metabolites, which may exhibit different variance patterns. Our study was also conducted in an agricultural community during the growing season, and it is possible that intermittent exposures related to nearby agricultural pesticide use may have caused higher variability in metabolite levels than would be observed in the general U.S. population. Furthermore, different exposure scenarios could result in different variability patterns. Before the phaseout of diazinon and chlorpyrifos in household pesticides (U.S. Environmental Protection Agency 2000, 2001), for example, children from homes with frequent indoor use of such pesticides may have shown more stable exposure levels. In a follow-up analysis, we intend to use additional data available to us to assess which potential exposure sources (e.g., diet, nearby agricultural pesticide use) best account for the metabolite variability patterns we present here.

### Conclusions
We found that though DAP metabolites in single or multiple spot samples are strongly correlated with levels in same-day 24-hr samples, children’s urinary DAP concentrations vary greatly from day to day, and use of spot samples to characterize a child’s cumulative weekly exposure results in a moderate degree of misclassification. Exposure misclassification resulting from urinary metabolite variability has the potential to bias measures of association between early childhood OP exposures and developmental outcomes in epidemiologic research toward the null hypothesis; such misclassification may account for the weak or null findings often reported to date between pediatric urinary DAP concentrations and child development (Bouchard et al. 2011; Marks et al. 2010). Additional research on variability in measures of nonpersistent compounds in children is needed to assure that exposure biomarkers are valid and that epidemiological studies have adequate power to detect health outcomes that may result from these exposures.

### References
Barr DB, Wilder LC, Caulliet SP, Gonzalez AJ, Needham LL, Pikula JL. 2005. Urinary creatinine concentrations in the

### Table 5. Sensitivity and specificitya for classifying children with high (top 20%) and elevated (top 40%) one-week average total DAP metabolite concentrationsb with one, two, or three urine samples.

<table>
<thead>
<tr>
<th>Type of sample</th>
<th>High (top 20%)</th>
<th>Elevated (top 40%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>Non-FMV samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One sample</td>
<td>0.52</td>
<td>0.89</td>
</tr>
<tr>
<td>Two samplesc</td>
<td>0.56</td>
<td>0.89</td>
</tr>
<tr>
<td>FMV samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One sample</td>
<td>0.58</td>
<td>0.90</td>
</tr>
<tr>
<td>Two samplesc</td>
<td>0.58</td>
<td>0.90</td>
</tr>
<tr>
<td>Any spot samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One sample</td>
<td>0.46</td>
<td>0.87</td>
</tr>
<tr>
<td>Two samplesc</td>
<td>0.60</td>
<td>0.90</td>
</tr>
<tr>
<td>Threesamples</td>
<td>0.64</td>
<td>0.91</td>
</tr>
</tbody>
</table>

aAverage sensitivity and specificity calculated based on 10 separate random samples of predictor spot samples.
bCalculations use creatinine-adjusted total DAP metabolite concentration (nmol/g creatinine) on the log10 scale. Pairs and trios consist of spot samples collected on different days (i.e., no pairs of same-day spots).


