Reviewer's report

**Title:** Correcting for the influence of sampling conditions on biomarkers of exposure to phenols and phthalates: a 2-step standardization method based on regression residuals.

**Version:** 1  **Date:** 6 February 2012

**Reviewer:** Anna Pollack

**Reviewer's report:**

- **Major Compulsory Revisions**

1. How can the authors conclude that the variation in concentrations of phenols and phthalates was due to sampling conditions rather than actual inter-individual variability? This case needs to be made more strongly because variation between individuals is to be expected due to differences in exposure to these chemicals.

2. What is the connection on p 5 lines 11-18 to the authors' work? How does the concept of a seasonal cutoff relate? Please clarify this point.

3. How do the authors actually correct for duration of storage before freezing when the Pelagie cohort does not have time of urine sampling? This question seems important. Did the authors consider using the Eden cohort to perform their standardization and then implement such corrections on the Pelagie cohort? This would be along the lines of a training and validation data set, where women from the Eden cohort were included in the training data set and the Pelagie cohort could be used to validate.

4. Were any sensitivity analyses run to test how the assumption of a 7:00 am sampling time for the Pelagie participants affected the results? (p 8 lines 9-11). Could the authors assess different sampling times to evaluate the sensitivity of their results to this finding? This is one of the main conclusions of the paper, that hour of sampling was associated with urinary concentrations of phthalate and phenol biomarkers, but for half the cohort, hour of sampling was unknown. Additional analysis and discussion on this point is needed.

5. The sampling conditions controlled for in step one should be described differently for the two cohorts included in the paper. Pelagie cannot effectively control for time spent at room temperature or hour of collection. Therefore, their step one standardization model depends only upon season, sampling day and gestational age at collection. This provides a fairly different amount of information on which to standardize, compared with Eden.

6. P 13 line 3-5, can this conclusion really be made for the Eden cohort, where time of collection is unknown?
7. Containers used to collect urine samples could have led to contamination of the samples and would be affected by length of time to processing. This is a fairly important issue given the goal of this paper. Do the authors know what type of containers were used in urine collection and could those leak BPA or other plasticizer components?

8. P 15 line 23-24, the standardization done in this paper does not account for measurement error per se. While unstandardized methods also do not account for measurement error, this sentence seems to mislead the reader into thinking that such standardizations take care of measurement error. Could the authors clarify this statement?

9. P 16 line 9-14, Although most work has focused on nutritional epidemiology where gold standards exist, there has been some work in the area without instrumental or gold standard variables to understand the effects of residual and unmeasured confounding, see Fewell Z, Davey SG, Sterne JA, 2009.

10. p 18 lines 16-18, The conclusion regarding fluctuation in levels by sampling time may not be supported by the entire data set. There is not yet evidence that the imputed value of 7:00 am for the Pelagie cohort was examined through sensitivity analyses or that this conclusion would hold in only the Eden cohort. Evidence of both should be made clear before such a strong conclusion can safely be made.

- Minor Essential Revisions
1. p. 5 line 1. Which participants would be excluded? This sentence is not entirely clear.
2. p. 5 line 4. Biomarker should not be plural. Please correct this and other instances.
3. p. 11 line 7-8 can you add what those correlations were for 2,5-DCP, MBP and MCPP?
4. Table 1. Date of urine sampling. This is somewhat unclear in the table. Perhaps a footnote or further explanation in the text could help clarify whether these are the end of each trimester or the midpoint within the trimester? I'm unsure how to interpret this row.
5. P. 13 line 7 can you indicate how strongly the standardized levels were correlated to the nonstandardized levels?

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**
I am funded at the NICHD under a grant to Dr. Enrique Schisterman from the Long Range Research Initiative of the American Chemistry Council. Other than this, I declare that I have no competing interests.