Author's response to reviews

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.

Authors:

Marion Mortamais (marion.mortamais@inserm.fr)
Cécile Chevrier (cecile.chevrier@univ-rennes1.fr)
Claire Philippat (claire.philippat@e.ujf-grenoble.fr)
Claire Petit (claire.petit@univ-rennes1.fr)
Antonia M Calafat (aic7@cdc.gov)
Xiaoyun Ye (xay5@cdc.gov)
Manori J Silva (zca2@cdc.gov)
Christian Brambilla (cbrambilla@chu-grenoble.fr)
Marinus J.C. Eijkemans (M.J.C.Eijkemans@umcutrecht.nl)
Marie-Aline Charles (marie-aline.charles@inserm.fr)
Sylvaine Cordier (sylvaine.cordier@univ-rennes1.fr1)
Rémy Slama (remy.slama@ujf-grenoble.fr)

Version: 2 Date: 6 March 2012

Author's response to reviews: see over
Dear Editor,

We thank you for offering us the opportunity to resubmit to *Environmental Health* our manuscript entitled:

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

The reviewer’s constructive comments have allowed us to clarify a number of points in the revised manuscript. In particular, we have made specific recommendations for future studies using biomarkers of exposure to non-persistent chemicals. A point by point answer to the reviewers’ comments is provided below.

Please also note that, after learning of discrepancies in the quality of phthalate metabolite analytical standards, which would affect the accuracy of the calculated metabolite concentrations, the CDC laboratory (Atlanta, in charge of the biomarkers assays in our study) discovered the analytical standards used were of inadequate purity. Consequently, the concentrations of two phthalate metabolites, monoethyl phthalate (MEP) and monobenzyl phthalate (MBzP), should be corrected (this applies to all MEP and MBzP assays performed by the CDCs in the past years). The correction factors are 0.66 (MEP) and 0.72 (MBzP) (A. Calafat, personal communication). In this new version, we have applied these correction factors.

We thank you in advance for considering our manuscript for publication in *Environmental Health*.

Sincerely,

Marion Mortamais

Rémy Slama, PhD

Senior Investigator

Head of the group of Environmental Epidemiology

Inserm U823, Grenoble
Reply to the Comments from the reviewers received Feb. 15th 2012

“Correcting for the influence of sampling conditions on biomarkers of exposure to phenols and phthalates: a 2-step standardization method based on regression residuals.”, Mortamais M et al.

Reviewers’ reports:

Reviewer I: Martine Vrijheid

This manuscript assesses the influence of sampling conditions on concentrations of phenol and phthalate biomarkers, and proposes a standardisation method to correct concentrations for sampling conditions. The reasons for this are well justified in the introduction: variations in sampling conditions can lead to exposure misclassification and few studies attempt to correct for such error. The methods and results are clearly described. The discussion tackles the most important concern about the use of standardised concentrations: the danger of over correction. My comments are minor. Some parts of the discussion can be improved to clarify the impact this type of correction will have on epidemiological dose-response analysis.

Minor essential revisions

I.1) 1. The relative changes between non-standardised and standardised concentrations are quite sizeable, but correlations are high (0.88-0.99). Can the authors comment on this in the discussion? What is the likely effect (if any) on dose-response results if the correlations are this high? Will the correction lead to narrower confidence intervals? or remove bias towards the null from random classical measurement error?

Reply: Exposure misclassification of a continuous or multicategorical variable can generally have impacts in various directions on dose-response relations and the associated confidence intervals in a specific study. If the error is purely random (e.g., if sampling conditions do not depend on exposure levels, which might be a strong assumption) then in expectation misclassification error due to sampling conditions may imply bias towards the null, at least if the exposure is dichotomized. Since this is not true if exposure has more than 2 levels (Dosimeci M, AJE, 1990)(which will be the case in analyses in which biomarkers levels are classified into 3 or more categories), we prefer not to draw too general conclusions regarding the impact of exposure misclassifications and of the impact of our approach used to correct it. Having said that, we agree that, in a study like ours where correlation coefficients between uncorrected and corrected values are above .80, we do not expect our correction to have a major impact on exposure misclassification (that is, reversing the direction of a significant trend). It still can affect p-values and dose-response relations to some extent. As an illustration, in a study of associations between phenol biomarkers levels and birthweight based on these data, the parameter associated with the raw concentration of 2,5 DCP (Dichlorophenol) corresponded to a decrease in birth weight by +32 g (95% CI, -120; 184) in the 2nd exposure tertile and -73 g (-206; 60) in the 3rd exposure tertile, compared to the lowest tertile (p for trend, 0.16) while the corresponding parameters were -49 g (95% CI, -206; 108) and -152 g (-299; -5, p for trend, 0.03) when the biomarkers concentrations standardized with our approach were used (Philippat C et al., EHP, in press).

I.2) It would be interesting to see the correlation coefficients between the standardised and non-standardised concentrations in table 4 instead of in the supplemental material.

Reply: We agree that it makes sense to report the correlation coefficients between standardized and
non-standardised concentrations in Table 4, together with the % change. This has now been added as an extra column.

I.3) 3. From the discussion on bottom of page 15 (statistical approach) it is not clear which approach (standardized or non-standardised) has less sources of variability: this text mentions “additional sources of variability”, “variability which varies”, etc. Then it states that “regression models ... should change this change in variance into account.” Please clarify these parts. It is unclear what is recommended here.

Reply: We thank the reviewer for pointing at this part of the discussion that indeed required clarification. We have clarified the text as follows:

“In a further step, one can use the standardized biomarkers concentration to assess the relation between biomarkers levels and specific health outcomes assessed in the same population [24]. Further developments of our approach that may be useful for this step would be to acknowledge for the variability in the regression coefficients corresponding to the effect of sampling conditions on biomarkers in the measurement model (Eq. A.1, see statistical appendix). Regression models in which the standardized concentrations are used as covariates should take this change in variance into account. Incidentally, it can be noted that using unstandardized (raw) levels in models not accounting for measurement error due to variability in sampling conditions will also impact on variance estimates and possibly bias; we believe that an approach like ours, aiming at making sources of measurement error explicit and correcting for them, is a step in the good direction.” (p.15)

I.4) 4. An important conclusion for future studies appears to be the collection of information on sampling conditions to make this type of analysis possible (many studies do not have this information). The discussion could be somewhat clarified on this: which conditions are most important to collect? Only sampling hour only? Would standardisation for only this variable give similar results to the full standardisation performed here?

Reply: We have shown that sampling hour, duration of urine storage at room temperature and gestational age can influence the concentration for some compounds. We now make it clearer in the discussion that future studies should collect information on all sampling conditions to allow a study of their impact as ours and to enable standardization whenever required.

Regarding the impact of sampling hour only, Table R1 below shows the results of standardized concentrations of MBP and MCPP for all sampling conditions, and compares it to the correlation coefficient obtained after standardization for sampling hour only. While MBP and MCPP levels are strongly influenced by the sampling hour (Table 2), the standardization for sampling hour does not explain alone the variation found between concentrations and non-standardized concentrations for all sampling conditions. For example, the correlation coefficient between standardized and non-standardized concentrations of MBP is .98 when only sampling hour is taken into account into the standardization and .96 when the other sampling conditions are additionally taken into account.

As suggested by the reviewer, we have indicated the following recommendation in the discussion section (p15):

“We chose to standardize each biomarker level on all sampling conditions simultaneously, but in future studies authors may prefer to standardize only for those sampling conditions that turn out to be associated with the considered biomarker with a p value below a given level (say, p<0.2).”

<table>
<thead>
<tr>
<th>Compound</th>
<th>n</th>
<th>Median</th>
<th>25&lt;sup&gt;th&lt;/sup&gt;</th>
<th>75&lt;sup&gt;th&lt;/sup&gt;</th>
<th>Median</th>
<th>25&lt;sup&gt;th&lt;/sup&gt;</th>
<th>75&lt;sup&gt;th&lt;/sup&gt;</th>
<th>Median</th>
<th>25&lt;sup&gt;th&lt;/sup&gt;</th>
<th>75&lt;sup&gt;th&lt;/sup&gt;</th>
<th>(2-1)&lt;sup&gt;e&lt;/sup&gt;</th>
<th>(3-1)</th>
<th>Pearson correlation coefficient between 2-1</th>
<th>Pearson correlation coefficient between 3-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phthalates metabolites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBP</td>
<td>287</td>
<td>48.1</td>
<td>28.9</td>
<td>86.7</td>
<td>58.1</td>
<td>34.3</td>
<td>105.6</td>
<td>50.2</td>
<td>29.3</td>
<td>85.4</td>
<td>+ 21%</td>
<td>+4%</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>MCPP</td>
<td>287</td>
<td>2.20</td>
<td>1.30</td>
<td>4.30</td>
<td>3.15</td>
<td>1.69</td>
<td>5.54</td>
<td>2.12</td>
<td>1.22</td>
<td>4.10</td>
<td>+ 43%</td>
<td>-4%</td>
<td>0.94</td>
<td>0.98</td>
</tr>
<tr>
<td>Phenols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,4-DCP</td>
<td>191</td>
<td>0.90</td>
<td>0.50</td>
<td>1.50</td>
<td>0.77</td>
<td>0.49</td>
<td>1.59</td>
<td>0.90</td>
<td>0.50</td>
<td>1.51</td>
<td>-14%</td>
<td>0%</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>Sum of Parabens (µmol/l)</td>
<td>191</td>
<td>0.86</td>
<td>0.27</td>
<td>2.45</td>
<td>0.95</td>
<td>0.29</td>
<td>2.32</td>
<td>0.88</td>
<td>0.26</td>
<td>2.33</td>
<td>+10%</td>
<td>+2%</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>BPA</td>
<td>191</td>
<td>2.70</td>
<td>1.70</td>
<td>5.50</td>
<td>3.11</td>
<td>1.92</td>
<td>6.11</td>
<td>2.92</td>
<td>1.73</td>
<td>5.40</td>
<td>+15%</td>
<td>+8%</td>
<td>0.95</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<sup>a</sup> For sum of parabens, concentrations are reported in µmol/l.

<sup>b</sup> Results corrected for over-representation of cases according to Richardson et al. (2007) and standardized for sampling conditions (excluding creatinine) using formula (1) in the methods section.

<sup>e</sup> Relative changes in the median concentration of urinary biomarkers between the estimates of column (5) and (2).
I.5) Results: text page 11 refers to “before 10:30” as sampling time but in the tables the cut-off is set at 10:00.

Reply: We indeed indicate in the text: “For Eden cohort, 95% of urine samples were collected before 10:30 AM”. However in Tables, the highest cut-off value did not correspond to the 95th centile but to 10:00 AM in order to have a sample size reasonably large in the highest category.

I.6) Tables: naming of the phthalates is different in Table 2 (metabolites) and Table 4 (parents).

Reply: We have now replaced parent compounds by their metabolites in Table 4.
Reviewer II: Anna Pollack

- Major Compulsory Revisions

**II.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.

Reply: We agree that between-subject variations in exposure levels are probably strongly driven by differences in exposure levels, in addition to variations due to sampling conditions and individual parameters associated with metabolizing capacity. However, our aim is to study the association between hour of urine sampling (and other sampling conditions) and biomarkers level, an association that can be studied without knowing the individual exposure levels provided that these exposure levels do not confound the association between sampling conditions and biomarkers levels. Therefore, our approach is valid under the assumption of a lack of association between exposure and sampling conditions (conditional on all individual parameters controlled for, such as BMI, SES or smoking). This is made explicit in the discussion:

> “We assumed that adjustment for individual characteristics such as age, occupation or smoking, made women with different sampling conditions more comparable. However, this approach might be limited by the existence of unmeasured lifestyle or occupational factors simultaneously associated with exposure and sampling conditions. For instance, if women who collected a urine sample early in the morning used more phthalate-containing cosmetics than those who provided a urine sample later in the day, we might attribute to variations in sampling hour differences actually due to real exposure contrasts. Time since last exposure (and amount of exposure) are also parameters likely to influence biomarkers levels. These were not available in our study; their assessment is challenging in observational studies not known by study participants. Moreover, time since last exposure is likely to be shorter for subjects frequently exposed to these compounds (and hence also probably more highly exposed to these compounds), so that standardization for these parameters might artificially decrease the between-subject contrasts in exposure.” (p 16)

This approach is similar to an epidemiological study on, say, vicinity to high voltage lines and children leukaemia, in which major determinants of childhood leukaemia (many of which are simply unknown) would not be controlled for. It can be noted that most (if not all) former studies on the influence of sampling hour on biomarkers levels also had no information on actual exposures (e.g., Calafat et al, EHP, 2008, based on the cross-sectional NHANES 2003-2004 population, Ye, EHP, 2011), which is very challenging given the number of exposure sources for most compounds considered, and the number of compounds studied.

**II.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.

Reply: The study by Wang et al on the association between the serum concentration of 25-hydroxyvitamin D and cancer risk relied on blood samples collected at various seasons. In this study (Wang et al.), the sampling condition possibly blurring exposure levels is season. We quote this attempt to correct for sampling conditions (here, by choosing season-specific cut-offs in biomarkers levels) as an early example of a study taking sampling conditions into account. We made the link to our study clearer:
"In purely descriptive studies, assayed biomarker levels are often left untransformed. When studying the impact of biomarker levels on health outcome, adjusting for sampling conditions influencing biomarker levels is sometimes performed. This approach may not be efficient because adjusting for sampling conditions in a regression model aiming at characterizing the effect of exposure on disease risk is unlikely to correct for the effect of sampling conditions on biomarker levels. As an illustration, in a study aiming at characterizing the association between serum concentration of 25-hydroxyvitamin D and cancer risk where between-subject differences in season of collection of blood sample existed, Wang et al. [19] considered several ways to handle differences in this sampling condition influencing the biomarker level. They have shown that, because of seasonal variations in 25-hydroxyvitamin D, choosing season-specific cut-offs to categorize the levels of this biomarker was a more efficient approach than adjusting for the date of sampling condition in a regression model where cancer occurrence was the dependent variable. Choosing season-specific cut-offs for categories of biomarker levels fluctuating with season is, in terms of identifying the group with the highest estimated exposure, equivalent to correcting biomarker levels by a value depending on the season of sampling. This approach has the advantage of being applicable independently of any information on health outcome, e.g., in descriptive (biomonitoring) studies.” (p 4)

II.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.

Reply: For Pelagie cohort, the storage time was determined as the number of hours between mailing of sample by post and receipt at the study center, assuming that mailing was done at 10:00AM and that the urine sample was received at 10:00 AM. Not knowing the exact sampling and reception hours will induce some error, but since for Pélagie cohort at least 2 days elapsed between sampling collection and reception in the research group, we assumed that the error of a few hours due to lack of knowledge of the exact hours could be neglected. We now indicate that the duration of storage before freezing was assessed “(in multiples of 24 hours for Pélagie cohort where sampling hour was unknown)”

The suggestion of using one of our cohort as a training dataset and the second as a validation dataset is appealing and would in principle be very relevant, but the differences between the 2 cohorts make this hard to implement. In particular, the fact that the main factor influencing biomarkers levels in Eden cohort is hour of sampling, which is not available in Pelagie cohort, does not allow replicating this result (which is however expected from what is known about the temporal variability of exposures and the short half-life of the studied compounds).

II.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.

Reply: We agree that the impact of sampling hour is a central result of our paper, but this was expected from the temporal variability of exposures and the short half-life of the studied compounds. We believe the main contribution of our manuscript relates to the approach we propose to standardize for sampling conditions. The additional analyses that we performed show that assuming another sampling hour for
Pelagie subjects would affect the average value of phthalate metabolites, by adding a similar constant value for all Pélagie subjects; it would very little affect the correlation between standardized and non-standardized biomarker levels (see Table R2 and reply to II.10 below). Effects regarding phenols would not be affected as these only rely on Eden cohort.

II.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.

Reply: For Pélagie cohort, time that urine sample spent at room temperature is available and was assessed as explained in response II.3). Moreover, when it was used as an adjustment factors, sampling time was assumed to be 7.00 AM for women of Pélagie, in order to include the entire population in the analyzes of the step 1 aiming at determine influence of the others sampling conditions (creatinine, gestational age, day of sampling, season of sampling, and duration of urine storage at room temperature) on biomarkers of exposure to phenols and phthalates. We recall that “models describing the influence of sampling time on biomarkers concentrations were estimated excluding Pélagie subjects” p 7.

II.6) P 13 line 3-5, can this conclusion really be made for the Eden cohort, where time of collection is unknown?

Reply: The reviewer certainly does speak of Pélagie cohort, for which hour of sampling is not available. As indicated in the previous reply (II.5), models describing the influence of sampling time on biomarkers concentrations were estimated excluding Pélagie subjects. We now indicate that this result is based on Eden cohort only.

II.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?

Reply: Containers used were made out of polypropylene (not containing BPA)(model FP40VPS, manufactured by CEB, Angers , France) but we cannot exclude that some women used another container than the one originally planned without the study midwife noticing it. The very fact that such contamination cannot be excluded make it important to check for an impact of duration of storage before freezing, as a way to identify such a potential problem, and to correct for it using an approach like the one we suggest. We indicate:

“The increasing concentration of BPA with increasing duration of storage at room temperature was unexpected; it might be due to a leakage of BPA from the plastic containers (or their caps) used to collect urine samples, as might happen if some women had used polycarbonate containers instead of the polypropylene containers planned for the study. Our analysis allowed to identify this potential issue and the statistical approach used attempted to correct for any resulting error.” (bottom of p.13 and top of p.14)
II.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?

Reply: If we define measurement error in broad terms such as the difference between the “ideal” exposure variable and the exposure variable used, then variability in sampling conditions constitutes one source of measurement error, together with any other deviations from the protocol, analytical errors, differences between the biologically relevant exposure time-window and the one used... We therefore believe that our expression referring to “measurement error due to variability in sampling conditions” is correct.

II.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.

Reply: We thank the reviewer for pointing the interesting reference by Fewell et al. (AJE, 2007, there is no 2009 publication from these authors). It is a simulation study aiming at characterizing the amplitude of the bias due to measurement error in the explanatory variable or in confounders; we agree that this allows to describe and understand the effects of residual and unmeasured confounding, but the simulation approach used allows authors to know “reality” (i.e. the assumptions of their simulation model), and therefore it can be seen as a study with a gold standard. Moreover, the Fewell et al study describes the extent of the bias, but does not suggest a statistical approach to correct for measurement error. Therefore this study does not contradict what we mentioned. We made it clearer in our text that, when we indicate that there is very little literature on the topic that we tackled, we mean literature suggesting approaches to correct for measurement error:

“There is a vast body of literature on how to handle and try to correct measurement error in covariates or health outcomes [35]. However, it is focused on situations in which there is some knowledge either on the standard errors attached to the error-prone variables or on the misclassification rate, on situations in which validation data in which both true and error-prone variables have been assessed in a sub-populations, or in which instrumental variables are available. These situations do not correspond to ours, in which we do know and measure some factors causing measurement error (the sampling conditions), and empirically estimate the influence of these factors on the mismeasured concentrations.”

Further, we now quote the reference by Fewell Z et al (AJE, 2007):

“The 2-step standardization method based on regression residuals that we proposed constitutes a way to reduce undesirable variability in biomarkers urinary concentrations due to sampling conditions, and allow more relevant comparisons between subjects and possibly between studies. This source of variability can be seen as a source of measurement error in exposure, which may have impacts in studies of the association between biomarkers levels and health, by impacting the regression models estimates in either direction (Fewell et al., 2007) and/or confidence intervals.”
II.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.

Reply: We agree that the conclusion regarding sampling hour is only based on results from Eden cohort, in which sampling hour was accurately assessed and which corresponds to about 2/3 of our sample size. This is clearly stated throughout the manuscript. We do not expect that changing the imputed sampling hour for Pélagie subjects would have a large impact, as this value is the same for all subjects. More precisely, this will shift the whole distribution of biomarkers levels in a similar manner for all Pélagie subjects and thus impact the change in the median values of standardized vs. unstandardized concentrations, but we do not expect this to alter much the coefficient of correlation between standardized and non-standardised concentrations. As an illustration, we report relative changes and coefficients of correlation making the assumption that all Pélagie samples have been collected at 9:00 AM instead of 7:00 AM (see Table R2 below): as can be seen, although median standardized biomarker concentrations do change, concentrations standardized assuming different sampling hours are very strongly correlated (coefficient, 0.99).
**Table R2. Phthalates Metabolites and Phenols Urinary Concentrations Among Pregnant Women From Eden and Pélagie Cohorts, France, 2002-2006.**

<table>
<thead>
<tr>
<th>Compound</th>
<th>n</th>
<th>Median</th>
<th>25th</th>
<th>75th</th>
<th>25th</th>
<th>75th</th>
<th>Median</th>
<th>25th</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP</td>
<td>287</td>
<td>48.1</td>
<td>28.9</td>
<td>86.7</td>
<td>58.1</td>
<td>34.3</td>
<td>105.6</td>
<td>72.9</td>
<td>40.7</td>
</tr>
<tr>
<td>MCCP</td>
<td>287</td>
<td>2.20</td>
<td>1.30</td>
<td>4.30</td>
<td>3.15</td>
<td>1.69</td>
<td>5.54</td>
<td>3.71</td>
<td>1.94</td>
</tr>
</tbody>
</table>

- Correction/standardization applied to the biomarker concentrations
- Case-control sampling only (µg/l) (1)
- Correction for case-control sampling and standardization for sampling conditions excluding creatinine (µg/l) (2)
- (2-1)b
- Correction for case-control sampling and standardization for sampling conditions excluding creatinine a (µg/l). Sampling time was assumed to be 9:00 AM for Pélagie (3)
- (3-1)b
- Pearson correlation coefficient between 2-1
- Pearson correlation coefficient between 3-1
- Pearson correlation coefficient between 3-2

- Results corrected for over-representation of cases according to Richardson et al. (2007) and standardized for sampling conditions (excluding creatinine) using formula (1) in the methods section.
- Relative changes in the median concentration of urinary biomarkers between the estimates of column (5) and (2).
Minor Essential Revisions

II.11) 1. p. 5 line 1. Which participants would be excluded? This sentence is not entirely clear.

Reply: We have now clarified this sentence:

“Because excluding participants not strictly adhering to the sampling protocol might induce selection bias, (…)”

II.12) 2. p. 5 line 4. Biomarker should not be plural. Please correct this and other instances.

Reply: This has been corrected as suggested throughout the document.

II.13) 3. p. 11 line 7-8 can you add what those correlations were for 2,5-DCP, MBP and MCPP?

Reply: We have added this information.

II.14) 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.

Reply: In Table 1, we described median values of all sampling dates (a footnote exists to clarify that these are the median values), and the 5th and 95th percentiles. For example, Nov. 2005 means that half of the urine samples were collected after Nov 2005, as we also now indicate in a footnote.

II.15) 5. P. 13 line 7 can you indicate how strongly the standardized levels were correlated to the nonstandardized levels?

Reply: We have added p 12 “correlation coefficients ranged between 0.88 for MBzP to 0.99 for triclosan”
III. Editorial comments

III.1) We agree with the reviewers that further information is needed to evaluate whether the manuscript is suitable for our journal. We note that the Conclusions in the Abstract begins with a caveat, which may be appropriate, but not useful to the reader who wants to know what this study shows.

Reply: We deleted this sentence and replaced it by a more positive statement:

“Sampling conditions, and in particular sampling hour, should be systematically collected in biomarker-based studies, in particular when biomarker half-life is short.”

III.2) In fact, our concern with this manuscript is that several variables seem to play a role in regard to the precision of the exposure biomarkers, and some of them have been evaluated in your study, others not. A revision would have to generate more specific recommendations to the readers how your experience should translate into recommendations on how to design future sampling regimens and analyses. Also, to the extent possible, you should try to highlight the sources of variability that are the most important so that future efforts can concentrate on controlling those.

Reply: We have now clarified our recommendations in several parts of the document (conclusion of the abstract, discussion section). In the discussion, we now indicate:

“We chose to standardize each biomarker level on all sampling conditions simultaneously, but in future studies authors may prefer to standardize only for those sampling conditions that turn out to be associated with the considered biomarker with a p value below a given level (say, p<0.2).” (p. 15, 2 lines before end)

“For this reason, studies on exposure-response relations using an approach such as ours should also report the association between the uncorrected biomarker concentrations and the health outcome, in addition to the association relying on standardized biomarker concentrations [24]. In addition, information on sampling conditions such as those considered here (in addition, whenever relevant, to batch number, assay date, and information on any deviation from the planned protocol) should be collected in all studies so that their possible impact can be characterized and if required corrected for.” (p. 18, 5 lines before end)

III.3) We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. We realize that our response may not be as optimistic as you might have hoped, but we believe that a revision prepared along these lines would attract much greater interest among our readers. Should you decide not to pursue this possibility, please let us know, and we will retract the submission. Please also ensure that your revised manuscript conforms to the journal style (http://www.ehjournal.net/info/instructions/).

Reply: This has been done.

III.4) It is important that your files are correctly formatted. The line and page numbering
throughout the MS should be removed.

Reply: This has been done.

**III.5** On the title page the heading above the email should read Email addresses.

Reply: This has been done.

**III.6** On the Abstract page the colons should be removed and the text move below the headings.

Reply: This has been done.

**III.7** The final heading in the Abstract and in the main text should be plural i.e. Conclusions.

Reply: This has been done.

**III.8** After the key words remove the abstract and text word counts.

Reply: This has been done.

**III.9** Sequential citations of three or more should be listed as [3-5].

Reply: This has been done.

**III.10** The second heading in the text should read Methods and all italicized text, except where indicated in the References changed to normal.

Reply: This has been done.

**III.11** There should be no headings in the Discussion section.

Reply: This has been done.

**III.12** After the Conclusions insert the following headings in this order: List of Abbreviations, Competing interests, Authors’ contributions and Acknowledgements.
Reply: This has been done.

**III.13)** All other text between the Conclusions section and the References should be removed.

Reply: This has been done.

**III.14)** The disclaimer should be included in the Acknowledgements.

Reply: This has been done.

**III.15)** In the List of Abbreviations, format as abbreviation:term separating the pairs with semi-colons in sentence format.

Reply: This has been done.

**III.16)** The names in the Authors' contributions section should be as initials, not full names and the sentence "All authors read and approved the final manuscript" inserted at the end.

Reply: This has been done.

**III.17)** Only the first letter of the heading Reference should be capitalized.

Reply: This has been done.

**III.18)** In the References section, please remove the issue numbers. For the Tables, all horizontal lines should be visible. Please try to fit all tables on one page if possible.

Reply: The issue numbers have been removed in the References section and all horizontal lines have been added. We realize that our tables are large, but we believe that information contained therein is relevant (only Tables 1 and 2 do not fit into one page).

**III.19)** The statistical appendix should be a separate (additional) file.

Reply: The statistical appendix will now be submitted in a separate document.

We look forward to receiving your revised manuscript by 15 April 2012. If you imagine that it will take longer to prepare please give us some estimate of when we can expect it.