HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2022 April 07.

Published in final edited form as:

Epidemiology. 2019 September ; 30(5): 756–767. doi:10.1097/EDE.0000000000001056.

An Empirical Validation of the Within-subject Biospecimens Pooling Approach to Minimize Exposure Misclassification in Biomarker-based Studies

Céline Vernet[a], **Claire Philippat**[a], **Lydiane Agier**[a], **Antonia M. Calafat**[b], **Xiaoyun Ye**[b,†], **Sarah Lyon-Caen**[a], **Pierre Hainaut**[c], **Valérie Siroux**[a], **Enrique F. Schisterman**[d], **Rémy Slama**[a]

[a]Inserm, CNRS, Univ. Grenoble Alpes, Institute for Advanced Biosciences (IAB), U1209, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, Grenoble, France;

[b]Organic Analytical Toxicology Branch, Division of Laboratory Sciences, Centers for Disease Control and Prevention, Atlanta, GA;

[c]Inserm, CNRS, Univ. Grenoble Alpes, Institute for Advanced Biosciences (IAB), U1209, Team of Tumor Molecular Pathology and Biomarkers, Grenoble, France;

[d]Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD.

Abstract

Background: Within-subject biospecimens pooling can theoretically reduce bias in dose–response functions from biomarker-based studies when exposure assessment suffers from classical-type error. However, collecting many urine voids each day is cumbersome. We evaluated the empirical validity of a within-subject pooling approach and compared several options to avoid sampling each void.

Methods: In 16 pregnant women who collected a spot of each urine void over several nonconsecutive weeks, we compared concentrations of 10 phenols in daily, weekly, and pregnancy within-subject pools. We pooled either three or all daily samples. In a simulation study using these data, we quantified bias in dose–response functions when using one to 20 urine samples per subject to assess methylparaben (a compound with moderate within-subject variability) and bisphenol A (high variability) exposures.

Results: Correlations between exposure estimates from pools of all and of only three voids per day were above 0.80 for all time windows and compounds, except for benzophenone-3

Correspondence: Rémy Slama, Institute for Advanced Biosciences – Centre de Recherche INSERM-CNRS-Univ. Grenoble-Alpes U1209, UGA Site Santé, Allée des Alpes, 38700 La Tronche, France. remy.slama@univ-grenoble-alpes.fr.

The code of the simulation can be obtained from the last author.

[†]Dr. Xiaoyun Sherry Ye passed away in 2018.

The authors report no conflicts of interest.

The findings expressed in this article are the opinions of the authors and do not necessarily reflect the official position of the Centers for Disease Control and Prevention (CDC). Use of trade names is for identification only and does not imply endorsement by the CDC, the Public Health Service, or the US Department of Health and Human Services.

Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

and triclosan in the daily time window (correlations, 0.57–0.68). With one spot sample to assess pregnancy exposure, correlations were all below 0.74. Using only one biospecimen led to attenuation bias in the dose–response functions of 29% (methylparaben) and 69% (bisphenol A); four samples for methylparaben and 18 for bisphenol A decreased bias to 10%.

Conclusions: For nonpersistent chemicals, collecting and pooling three samples per day instead of all daily samples efficiently estimates exposures over a week or more. Collecting around 20 biospecimens can strongly limit attenuation bias for nonpersistent chemicals such as bisphenol A.

Keywords

Attenuation bias; Exposure assessment; Exposure biomarkers; Measurement error; Phenols; Pooling; Sampling design; Within-subject variability

Investigating the potential human health impact of environmental pollutants requires an accurate estimation of a proxy exposure over relevant time windows.[1,2] For chemicals with multiple or poorly characterized sources, given the high sensitivity of targeted biochemical assays, exposure biomarkers are the most frequently used option in human studies. However, despite its analytical accuracy, this approach may introduce strong exposure misclassification. Indeed, for chemicals whose biomarker concentrations display high (within-subject) temporal variability (e.g., phenols, phthalates, dialkyl phosphates),[3–5] relying on a single or a couple of biospecimens per subject provides a poor estimate of the average exposure over time. In the case of classical-type error, this is expected to lead to (sometimes strong) attenuation bias in dose–response relationships.[2,6] Classical-type measurement error occurs when individual's biomarker concentrations vary around the true value, which can be approximated by the mean of repeated measurements throughout the target time window.[7]

Increasing the number of biospecimens collected from each subject mitigates attenuation bias.[2,8] If several biospecimens are available in at least part of the study population, one can quantify biomarker concentrations in each biospecimen and use measurement error models to limit bias.[7,9,10] This approach increases analytical costs. One alternative consists in pooling the biospecimens within-subject, which benefits from the repeated exposure information collected for each subject, without increasing analytic costs because one pool is analyzed per subject. This so-called within-subject biospecimens pooling approach has been validated theoretically,[8] but its implementation in large-scale epidemiologic studies and for long exposure windows (e.g., the entire pregnancy) raises practical issues. In particular, collecting all daily urine samples may be cumbersome; consequently, evaluating the efficiency of downgraded approaches in which only a few daily voids are sampled would help designing efficient and feasible studies.

Our aims were (1) to evaluate the efficiency of a sampling design based on collecting three spot urine samples per day to approximate the average exposure over daily, weekly, and whole pregnancy exposure windows, compared with collecting all daily urine voids; and (2) to empirically investigate the effect of within-subject temporal variability in biomarkers concentrations in actual populations on bias in dose–response functions. We used select

phenols as examples of nonpersistent biomarkers covering a large range of within-subject variability.

METHODS

Overview

We relied on pregnant women recruited for the SEPAGES (Suivi de l'Exposition à la Pollution Atmosphérique durant la Grossesse et Effets sur la Santé; Assessment of air pollution exposure during pregnancy and effects on health) cohort feasibility study, who collected a spot sample of each urine void over three pregnancy weeks.[3,11] Urine was pooled within-subject using different approaches, from which we assessed the agreement between two estimates of daily, weekly, and pregnancy exposure: one relied on pools made from all daily samples for each subject, and one on pools using fewer samples (aim 1). From the same exposure dataset, we generated a fictitious study, assumed phenols impacted a health outcome, and evaluated the impact on the estimated dose–response function of increasing the number of biospecimens for exposure assessment, and of using an a posteriori disattenuation approach[8] (aim 2).

Biospecimens

Urine Collection—SEPAGES feasibility was approved by the appropriate ethical committees (CPP, Comité de Protection des Personnes Sud-Est; CNIL, Commission Nationale de l'Informatique et des Libertés; CCTIRS, Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé; ANSM, Agence Nationale de sécurité du Médicament et des produits de santé). All participants provided written informed consent for biologic measurements and data collection. The involvement of the Centers for Disease Control and Prevention (CDC) laboratory did not constitute engagement in human subjects research.

As detailed elsewhere,[3] 30 pregnant women living in Grenoble (France) urban area collected a spot sample of each urine void in polypropylene containers during three nonconsecutive weeks (median: 13, 23, and 32 gestational weeks) between July 2012 and July 2013.

Samples were kept in the participants' refrigerators until the study staff retrieved, aliquoted, and froze them at −80°C into polypropylene cryovials at Inserm Institute for Advanced Biosciences, Grenoble. Women recorded any missed voids. We quantified biomarkers in the 16 women with the smallest number of missed voids. Two women collected a sample of all their voids (no missing void, group A1), six more than 95% (group A2), and eight between 80% and 95% of their voids (group B).

Urine Pools—We pooled individual samples as detailed in Figure 1. For each subject, we prepared (i) within-subject daily pools (seven daily pools per subject per week) using an equal volume of all voids of each day (there were on average eight voids per day); (ii) within-subject weekly pools, obtained by pooling an equal volume of all daily pools of each of the 3 weeks; and (iii) a within-subject pregnancy pool, obtained by pooling all weekly pools. This ideal approach corresponds to protocol 1.

In a downgraded pooling approach (protocol 2), daily pools were prepared using three randomly selected samples (instead of all in protocol 1) from each subject: one from the morning (mid-night to 11:59 A.M.), one from the afternoon (12:00–5:59 P.M.), and one from the evening (6:00–11:59 P.M.). Weekly and pregnancy pools were prepared as in protocol 1 from these downgraded daily pools (Figure 1).

All samples were kept frozen at −80°C in 2 ml polypropylene cryovials until shipment on dry ice to the CDC (Atlanta, GA), where all biospecimens were stored (−70°C) until analysis.

Quantification of Phenol Biomarkers

We quantified total (free plus conjugated) concentrations of 2,4-dichlorophenol, 2,5-dichlorophenol, benzophenone-3, bisphenol A, bisphenol S, triclosan, butylparaben, methylparaben, ethylparaben, and propylparaben using online solid-phase extraction high-performance liquid chromatography-isotope dilution-tandem mass spectrometry.[12]

We quantified phenols in pregnancy pools of all 16 women. Because of cost limitations, we quantified phenols in weekly pools only in the eight women of groups A1 and A2 (group A); in daily pools of a whole week in the two women from group A1; and in only one random day in group A2 (20 daily pools in total).

Aim 1: Efficiency of a Downgraded Within-subject Pooling Protocol for Exposure Assessment

We compared the efficiency of the downgraded pooling approach (protocol 2), with that of the ideal pooling approach (protocol 1, our reference) to provide an estimate of exposure to the 10 considered phenols or their precursors over exposure windows of a day, a week and the whole pregnancy (corresponding to the average of the 3 follow-up weeks).

For the pregnancy exposure window, we also compared exposure estimates from the ideal approach to other downgraded approaches based on reliance on one to eight spot samples randomly selected from all available samples collected for each of the eight women of group A:

- Protocol 3, relying on a single random spot sample, consistent with the typical epidemiologic study published in recent years (with the difference that samples are generally not randomly selected);
- Protocols 4 and 5, relying on the averaged biomarker concentrations in three and eight random spot samples, respectively.

Nondetectable concentrations were replaced by instrumental readings and, for null instrumental readings, by the biomarker-specific non-null lowest instrumental reading divided by √2. Biomarker concentrations were ln-transformed.

We compared phenol concentrations averages (as well as creatinine and specific gravity measurements) between protocols using correlation coefficients, paired *t* tests, Cohen's Kappa coefficients, scatter plots, and Bland-Altman plots.[13,14]

Aim 2: Impact of Biomarker Variability on Dose–Response Estimates

Methods are detailed in eAppendix 1; http://links.lww.com/EDE/B550. Our approach parallels that described in a theoretical study in which biomarker concentrations were simulated,[8] with the difference that we relied here on phenol urinary concentrations from eight spot samples randomly selected from all available biospecimens collected throughout pregnancy in eight women (group A). A bootstrap approach was used to generate populations of 3,000 subjects with one to 20 biospecimens each. We quantified bias and statistical power of epidemiologic studies aiming at relating exposure to two phenol biomarkers to a continuous health outcome (assumed to correspond to child weight at 3 years). We chose two compounds, methylparaben (intraclass correlation coefficient, ICC = 0.85) and bisphenol A (ICC = 0.38), because of their contrasted pregnancy-specific ICCs in the studied population of eight women.[3] Exposure was assumed to be assessed from biomarker concentration in one random spot sample or in within-subject pools of an increasing number of biospecimens.

Bias was estimated as the difference in percent between the mean effect estimate (β) over 1,000 studies for the surrogates of exposure and the true effect (β_{true}), divided by the true effect. Statistical power was calculated as the fraction of the 1,000 studies with a *P* value for the association below 0.05.

We additionally reported a posteriori disattenuated effect estimates.[2,7,8] These estimates were obtained by dividing the regression coefficient associated to each compound by the ICC of the compound. We used two pregnancy-specific ICC: ICC_1, ICC corresponding to the true value for this specific population, and estimated in our study population of eight women,[3] and ICC_2, the average ICC from previous studies of pregnant women, namely 0.45 (methylparaben) and 0.20 (bisphenol A).[15–20] We analyzed data using STATA 12.1 (Stata Corp, College Station, TX).

RESULTS

Population

Each woman collected between three and 15 urine specimens per day (median, 7, 25th–75th centiles, 6–10) over 3 weeks (total number of samples per pregnancy: median, 160, 25th–75th centiles, 136–188). A median time of 8.9 weeks separated successive follow-up weeks (eTable 1; http://links.lww.com/EDE/B550).

Efficiency of a Downgraded Within-subject Pooling Protocol

Daily Exposure Window—In daily pools from protocol 1, detection frequencies were above 75% for all phenols except benzophenone-3 (45%). Using three voids per day (protocol 2), detection frequencies were similar for most compounds except triclosan (50%) and benzophenone-3 (30%, Table 1).

Biomarker daily averages were similar between protocols 1 and 2 for all biomarkers except bisphenol S and triclosan; for triclosan, there was a trend of underestimation of daily averages with protocol 2 compared with protocol 1 (*t* test *P* value < 0.001, eFigures 1–2;

http://links.lww.com/EDE/B550). Pearson correlations between ln-transformed biomarker concentrations from protocols 1 and 2 were highest for parabens ($r \geq 0.96$, Table 1) and above 0.80 for all compounds except benzophe-none-3 ($r = 0.57$) and triclosan ($r = 0.68$).

Weekly Exposure Window—Detection frequencies in weekly pools were similar to those observed for the daily pools, except for butylparaben (58% in weekly pools, 85% in daily pools, Tables 1, 2). Protocol 2 weekly concentration averages were similar to those of protocol 1, except for 2,5-dichlorophenol, propylparaben, and bisphenol A, with median concentrations tending to be higher in protocol 2, compared with protocol 1 (eFigures 3–4; http://links.lww.com/EDE/B550). Ln-transformed weekly biomarker concentrations were highly correlated between both protocols ($r > 0.8$), the lowest correlation being for triclosan ($r = 0.81$) and the highest for three parabens and 2,5-dichlorophenol ($r \geq 0.98$, Table 2).

Pregnancy Exposure Window—In pregnancy pools, detection frequencies (median, 25th–75th centiles) were similar between protocols 1 (97%, 83%–100%) and 2 (97%, 90%–100%), and generally lower for protocol 3 (69%, 50%–88%), except for benzophenone-3 (63% [protocol 3] vs. 31% [protocols 1 and 2], Table 3).

For all biomarkers, protocols 1 and 2 pregnancy averages (n = 16 women) were in close agreement ($r \geq 0.86$). Regarding protocols 3–5 (n = 8), correlations with estimates from protocol 1 increased with the number of spot biospecimens used to assess pregnancy exposure: depending on the compounds, correlations ranged from −0.67 to 0.74 for protocol 3 (one biospecimen during pregnancy), from 0.60 to 0.92 for protocol 4 (three biospecimens), and from 0.68 to 0.98 for protocol 5 (eight biospecimens, Table 3, Figure 2). For protocols 4–5, correlations with protocol 1 were above 0.80 for all biomarkers but bisphenols and triclosan. Scatter plots and Bland-Altman plots suggested underestimation of pregnancy exposure when using protocols 3–5, compared with protocol 1 (eFigures 5–6; http://links.lww.com/EDE/B550).

Impact of Biomarker Variability on Dose–Response Estimates

One Biospecimen for Exposure Assessment—When using one random spot sample per subject to assess exposure, the average effect estimate for methylparaben was −71 g (95% confidence interval [CI], −101, −40), corresponding to an attenuation bias of 29% compared with the true effect of −100 g. Power was 99% (eTable 2; http://links.lww.com/EDE/B550).

For bisphenol A, relying on a single spot sample led to an average effect estimate of −31 g (95% CI, −76, 16; attenuation bias, 69%). Power was 27% (eTable 3; http://links.lww.com/EDE/B550).

A posteriori disattenuation using study-specific ICCs (ICC_1) reduced the attenuation bias to 16% (methylparaben) and 19% (bisphenol A). By contrast, disattenuation applied with an average value of the biomarker-specific ICC from external studies (ICC_2) overcorrected the effect estimate for both biomarkers: bias was +58% for methylparaben (compared with −29% without correction) and +54% for bisphenol A (compared with −69%, eTables 2–3; http://links.lww.com/EDE/B550).

For both compounds, type 1 error rate did not increase (5%) when no effect of the true exposure was assumed (data not shown).

Increasing the Number of Biospecimens—Four (methylparaben) and 18 (bisphenol A) samples were required to limit bias below 10% (Figure 3). When we applied disattenuation using study-specific ICCs, the numbers of samples required were two (methylparaben) and three (bisphenol A).

DISCUSSION

Assessing exposure biomarkers from one spot biospecimen per subject provides an error-prone estimate of exposure to chemicals with strong within-subject variability. Within-subject pooling of biospecimens is an efficient way to estimate exposure. We compared four downgraded protocols with an ideal approach to provide insight about whether the within-subject biospecimens collection could be simplified without increasing error. Within-subject pooling of three samples per day was almost as efficient as collecting all daily samples to assess exposure averages over short (1 week) to longer time periods (the whole pregnancy) for phenols with relatively short elimination half-lives. This was also the case for shorter (daily) time windows, except for benzophenone-3 and triclosan. When the entire pregnancy was the targeted exposure window, collecting three to eight random spot biospecimens was almost as efficient as collecting all or three samples per day for several weeks, but not for all chemicals.

Exposure misclassification from reliance on spot biospecimens entails bias in dose–response functions; we provided an empirical estimation of the amplitude of the corresponding attenuation bias, which was strong for compounds with high within-subject variability such as bisphenol A. Increasing the number of biospecimens collected per subject reduced the attenuation bias and increased statistical power. Applying the ICC-based a posteriori disattenuation method[8] corrected part of the attenuation bias. The same approach using external ICCs derived from the literature increased the amplitude of the bias, showing the strong sensitivity of this method to the ICCs used.

Study Assumptions and Limitations

By comparing exposure estimates in a small group of women (n = 8), we may have reduced the variability of exposure biomarkers, possibly increasing measured correlations. However, the number of biospecimens analyzed (n = 124) to make these comparisons was large. Missed voids may have artificially increased correlations between the two within-subject pooling approaches, by lowering the number of specimens in the ideal pooling approach. We limited this problem by selecting women with few missed voids.

There are several ways to combine biospecimens within-subject. We created equal-volume pools, the simplest option in practice. Alternatives include pooling proportionally to the total volume of each void (a rather cumbersome approach), or taking into account the dilution of each urine sample (e.g., pooling volumes proportional to urine dilution). All pooling protocols were similarly affected by this choice, which might not strongly impact the agreement between the compared pooling protocols. The present study was restricted to

a specific population and selected chemicals. Hence, generalization of our results should be considered with great caution, as we discussed elsewhere.[3]

In the simulation, we assumed that measurement error was of classical type, a reasonable assumption for exposure biomarkers.[7,21]

Assessing Exposure Over Time Windows of Various Lengths

For an exposure window of several weeks (typically the whole pregnancy), biomarker concentrations from a single random spot sample were, in general, in poor agreement with pregnancy exposure averages from the ideal protocol 1, confirming that relying on a single random spot sample does not always accurately represent the pregnancy average. Increasing the number of biospecimens per subject improved the agreement for most of the studied chemicals, with fair agreement with protocol 1 pregnancy exposure averages when relying on three to eight random biospecimens, except for triclosan and bisphenols. However, although the exposure ranking was preserved (as seen from the correlation coefficients), when using three to eight spot samples, the pregnancy average concentration was generally not perfectly estimated. This may not be an issue when one is interested in estimating the slope of a dose–response function assumed to be linear, but becomes one when dose–response functions are not linear, or for biomonitoring studies. For chemicals with variability such as that of ethylparaben, triclosan, and bisphenols, relying on eight random spot samples to assess pregnancy exposure may not be enough. Our previous study in the same population reported high within-subject variability for these compounds.[3] In contrast to protocols relying on random spot samples, agreement with the ideal approach was high for all biomarkers when repeatedly collecting three daily samples, which confirmed that this approach is efficient to assess exposure over both short and long exposure windows, even for some of the chemicals with highly variable concentrations.

Efficiently characterizing the average exposure over a week could be achieved collecting three biospecimens per day that week. When it came to characterizing exposure over a day, the approach was still efficient, except for benzophenone-3 and triclosan (correlations in the 0.5–0.7 range). However, caution is required in interpreting these results, as these biomarkers had the lowest detection frequencies.

For a few chemicals (e.g., bisphenol A and triclosan for the daily window; 2,5-dichlorophenol, propylparaben, and benzophenone-3 for the weekly window), exposure averages differed between protocols 1 and 2, but exposure rankings were preserved for all compounds except triclosan in the daily window. Results for triclosan are quite consistent across exposure windows; be it for exposure ranking or dose–response function estimation, collecting half a dozen of biospecimens in the exposure window of interest may not be sufficient to assess exposure.

Bias in Dose–Response Functions

Attenuation in regression analyses is well known in the context of classical-type error.[7,8,21] Using real data, we quantified the attenuation bias in regression parameters when a single error-prone biomarker measurement is used as surrogate of the true underlying exposure. Our findings confirm theoretical results according to which bias is related to ICC.[2,7,8]

Perrier et al.[8] who, using simulated exposure data, reported an attenuation bias of 80% for highly variable biomarker concentrations (ICC = 0.2), and of 40% for chemicals with less variable concentrations (ICC = 0.6).

Without a posteriori disattenuation, four samples were required for methylparaben and 18 for bisphenol A to limit bias below 10%, compared with six and 35 samples, respectively, in Perrier et al.[8] Hence, we observed an attenuation bias of lower magnitude and a smaller number of biospecimens required to efficiently reduce bias. This may be due to the relatively high ICC values in our study population, compared with those assumed by Perrier et al.[8] Our overall assessment is that a few (three to five) biospecimens over a specific time window are required for chemical biomarker with a relatively low within-subject variability, whereas for highly variable chemical biomarkers, at least one or two dozen biospecimens are needed.

A posteriori disattenuation, a simple technique to limit the impact of classical-type measurement error on dose–response function estimates, partly reduced the attenuation bias when using the ICCs observed in our population, which differed from the perfect correction observed by Perrier et al.[8] Perrier et al.[8] simulated data using a predefined ICC, while we estimated ICCs from a small sample size (n = 8 women), which may have reduced the precision of the ICC estimates. Using ICCs extracted from the literature overcorrected regression estimates (i.e., created a bias in the opposite direction with greater magnitude). Discrepancies in the temporality of urine collection between studies may partly explain the lack of validity of external ICCs, because ICCs depend on the biospecimens sampling frame.[3] Also, exposure sources, pathways, and toxicokinetics of chemicals may vary between populations, resulting in different ICCs. This underlines the relevance of estimating variability for each study, e.g., by analyzing multiple spot biospecimens from a subsample of the study population to correct bias using a posteriori disattenuation, even though within-subject pooling is used. When population-specific ICCs are not available, employing external ICCs to correct estimates should be done with great caution, if at all.[7,22]

Within-Subject Pooling Approach

The downgraded within-subject pooling approach allows the investigation of short (days, weeks) or long (trimesters of pregnancy) exposure windows for the target biomarkers, despite limited efficiency for benzophenone-3 and triclosan in the shortest time windows (day) for the present study population. Overall, such an approach, without being too cumbersome for study participants, permits to combine information from many biospecimens to estimate exposure averages and reduce attenuation bias in effect estimates in dose–response functions, without increasing analytical costs because a single pooled sample per woman is analyzed for a target exposure window.[8,23] We have applied this approach in 479 pregnant women from SEPAGES couple–child cohort recruited in 2014–2017 in Grenoble area,[24] and in a subgroup of HELIX exposome project participants,[25] showing the feasibility of its implementation.

We assumed that pooling samples did not entail any error. Pooling error may exist because of technical differences (e.g., technician-related variability, instruments precision); physical conditions (e.g., ambient temperature, thawing duration)[26]; or of lack of consideration of

urinary dilution in the pooling strategy. In our study, a single technician pooled samples, limiting error due to biospecimens manipulation.

Collecting and pooling three daily urine specimens over toxicologically relevant exposure windows has the advantage of being less cumbersome than collecting all urine voids. Compared with collecting a single spot biospecimen per subject, the logistic burden and overall study costs are increased, which may affect sample size. Some subjects may be reluctant to repeatedly collect biospecimens, but these should not be excluded because unbalanced designs (with varying number of biospecimens per subject) can give acceptable estimates of dose–response functions, despite a slightly higher bias in effect estimates.[8] If collecting repeated samples is only possible among a few subjects, then this should still be undertaken, to provide an internal estimate of ICCs, which could be used to apply a posteriori disattenuation to estimates, at least as a sensitivity analysis.

CONCLUSIONS

Whatever its sensitivity and accuracy, quantification of nonpersistent chemicals in a spot biospecimen is expected to provide a poor estimate of exposure over time windows of a day or more. One relevant alternative is repeated collection of within-subject biospecimens without pooling or a within-subject biospecimens pooling approach.[8] We demonstrated here for a large family of chemicals and range of within-subject variability in biomarker concentrations that a sampling approach relying on repeated within-subject pooling of three daily spot samples over a target exposure window (days, weeks, whole pregnancy) can accurately estimate exposure averages over this time window, without increasing analytical costs or being excessively cumbersome. Not pooling biospecimens and measuring exposure biomarkers in each biospecimen allows reliance on measurement error models,[7,8] further limiting bias but for a larger assay cost.

The within-subject biospecimens pooling approach appears to be a cost-efficient solution to minimize exposure misclassification related to the temporal variability of nonpersistent exposure biomarkers. It may also be considered for nonpersistent effect biomarkers, such as hormonal levels or seminal parameters.[27] While repeating (and possibly averaging) measurements within each observation unit is a basic metrological principle in other areas of health and environmental sciences (e.g., assessment of blood pressure or air pollution exposure), it is currently seldom applied to biomarker-based studies. Within-subject collection of repeated samples, with or without pooling, could allow epidemiologists making the best of the “biomarker revolution”[28] for biomarkers with high within-subject variability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Mrs. L. Borges, clinical research assistant, and the staff from Grenoble Center for Clinical Investigation (CIC): Professor J.-L. Cracowski, Dr. E. Hodaj, Mrs. D. Abry, Mrs. A. Tournier, Mrs. J. Quentin, and Mr. N. Gonnet. The support of Dr. M. Althuser, Dr. F. Camus-Chauvet, Dr. D. Marchal André, Dr. X. Morin, Dr. P. Rivoire,

Mrs. A. Royannais, Dr. C. Tomasella, Dr. T. Tomasella, Mr. P. Viossat, Mrs. E. Volpi, Mrs. S. Rey, Professor P. Hoffmann and clinicians from Grenoble University Hospital in the recruitment of the study volunteers. We thank Mrs. F. Perrier and Mr. A. Ye for their help with biostatistical aspects. We also acknowledge the technical assistance of Dr. P. Dwivedi, Ms. X. Zhou, Ms. T. Jia, Ms. P. Olive, and Mr. T. Powell (Centers for Disease Control and Prevention, Atlanta, GA) in measuring the urinary concentrations of phenol biomarkers, and creatinine. Finally, we sincerely thank the participants of the SEPAGES feasibility study.

Supported by the European Research Council (ERC consolidator grant N°311765-E-DOHaD, Principal investigator: Rémy Slama); ANR Shalcoh (ANR-14-CE21-0007); Fonds Agir Pour les Maladies Chroniques 2011 (APMC, CDMR R13076CC); AGIRàdom. E.F.S. is supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health. C.V. benefited of a doctoral grant from University Grenoble Alpes.

REFERENCES

1. Calafat AM, Longnecker MP, Koch HM, et al. Optimal exposure biomarkers for nonpersistent chemicals in environmental epidemiology. Environ Health Perspect. 2015;123:A166–A168. [PubMed: 26132373]
2. Rappaport SM, Symanski E, Yager JW, Kupper LL. The relationship between environmental monitoring and biological markers in exposure assessment. Environ Health Perspect. 1995;103(suppl 3):49–53.
3. Vernet C, Philippat C, Calafat AM, et al. Within-day, between-day, and between-week variability of urinary concentrations of phenol biomarkers in pregnant women. Environ Health Perspect. 2018;126: 037005. [PubMed: 29553460]
4. Casas M, Basagaña X, Sakhi AK, et al. Variability of urinary concentrations of non-persistent chemicals in pregnant women and school-aged children. Environ Int. 2018;121(pt 1):561–573. [PubMed: 30300814]
5. Romano ME, Hawley NL, Eliot M, et al. Variability and predictors of urinary concentrations of organophosphate flame retardant metabolites among pregnant women in Rhode Island. Environ Health. 2017; 16:40. [PubMed: 28399857]
6. Brunekreef B, Noy D, Clausing P. Variability of exposure measurements in environmental epidemiology. Am J Epidemiol. 1987;125: 892–898. [PubMed: 3565364]
7. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 2006.
8. Perrier F, Giorgis-Allemand L, Slama R, Philippat C. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. Epidemiology. 2016;27:378–388. [PubMed: 27035688]
9. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. J Am Stat Assoc. 1994;89:1314–1328.
10. Hardin JW, Schmiediche H. The regression-calibration method for fitting generalized linear models with additive measurement error. The Stata Journal. 2003;3:361–372.
11. Ouidir M, Giorgis-Allemand L, Lyon-Caen S, et al. Estimation of exposure to atmospheric pollutants during pregnancy integrating space-time activity and indoor air levels: does it make a difference? Environ Int. 2015;84:161–173. [PubMed: 26300245]
12. Zhou X, Kramer JP, Calafat AM, Ye X. Automated on-line column-switching high performance liquid chromatography isotope dilution tandem mass spectrometry method for the quantification of bisphenol A, bisphenol F, bisphenol S, and 11 other phenols in urine. J Chromatogr B Analyt Technol Biomed Life Sci. 2014;944:152–156.
13. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135–160. [PubMed: 10501650]
14. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat. 2007;17: 571–582. [PubMed: 17613642]
15. Braun JM, Kalkbrenner AE, Calafat AM, et al. Variability and predictors of urinary bisphenol A concentrations during pregnancy. Environ Health Perspect. 2011;119:131–137. [PubMed: 21205581]

16. Braun JM, Smith KW, Williams PL, et al. Variability of urinary phthalate metabolite and bisphenol A concentrations before and during pregnancy. Environ Health Perspect. 2012;120:739–745. [PubMed: 22262702]
17. Jusko TA, Shaw PA, Snijder CA, et al. Reproducibility of urinary bisphenol A concentrations measured during pregnancy in the Generation R Study. J Expo Sci Environ Epidemiol. 2014;24:532–536. [PubMed: 24736100]
18. Meeker JD, Cantonwine DE, Rivera-González LO, et al. Distribution, variability, and predictors of urinary concentrations of phenols and parabens among pregnant women in Puerto Rico. Environ Sci Technol. 2013;47:3439–3447. [PubMed: 23469879]
19. Philippat C, Wolff MS, Calafat AM, et al. Prenatal exposure to environmental phenols: concentrations in amniotic fluid and variability in urinary concentrations during pregnancy. Environ Health Perspect. 2013;121:1225–1231. [PubMed: 23942273]
20. Smith KW, Braun JM, Williams PL, et al. Predictors and variability of urinary paraben concentrations in men and women, including before and during pregnancy. Environ Health Perspect. 2012;120: 1538–1543. [PubMed: 22721761]
21. Fuller WA. Measurement Error Models. 1st ed. Hoboken, NJ: Wiley-Interscience; 2006.
22. Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Stat Med. 1994;13:1265–1282. [PubMed: 7973207]
23. Schisterman EF, Vexler A. To pool or not to pool, from whether to when: applications of pooling to biospecimens subject to a limit of detection. Paediatr Perinat Epidemiol. 2008;22:486–496. [PubMed: 18782255]
24. Slama R, Vernet C, Nassan FL, Hauser R, Philippat C. Characterizing the effect of endocrine disruptors on human health: the role of epidemiological cohorts. C R Biol. 2017;340:421–431. [PubMed: 28843646]
25. Vrijheid M, Slama R, Robinson O, et al. The human early-life exposome (HELIX): project rationale and design. Environ Health Perspect. 2014;122:535–544. [PubMed: 24610234]
26. Schisterman EF, Vexler A, Mumford SL, Perkins NJ. Hybrid pooled-unpooled design for cost-efficient measurement of biomarkers. Stat Med. 2010;29:597–613. [PubMed: 20049693]
27. Tielemans E, Heederik D, Burdorf A, Loomis D, Habbema DF. Intraindividual variability and redundancy of semen parameters. Epidemiology. 1997;8:99–103. [PubMed: 9116104]
28. Schisterman EF, Albert PS. The biomarker revolution. Stat Med. 2012;31:2513–2515. [PubMed: 22969023]

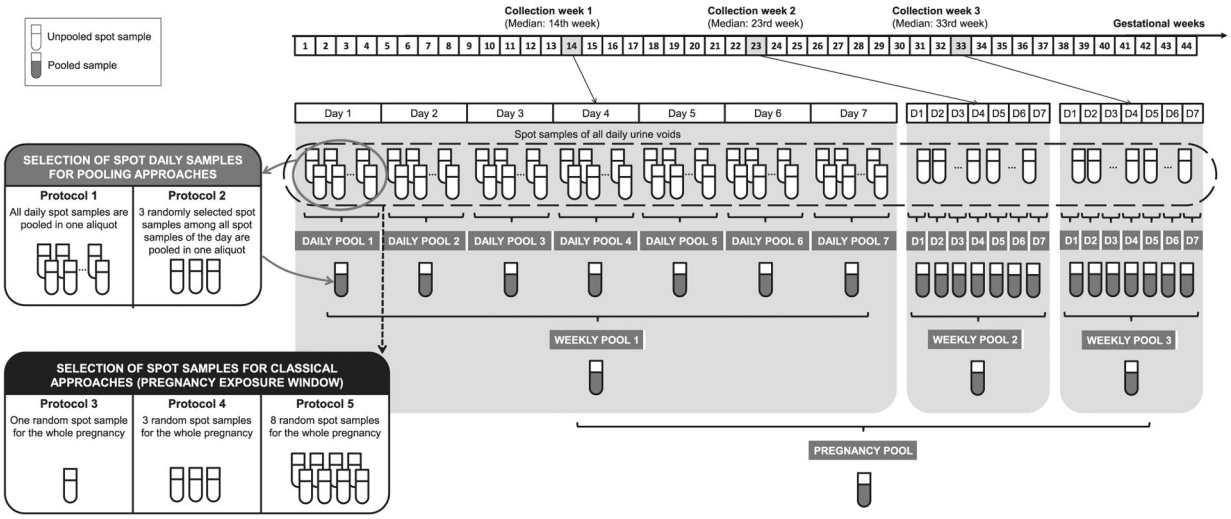

FIGURE 1.
Study design (n = 16 pregnant women from SEPAGES cohort feasibility study).

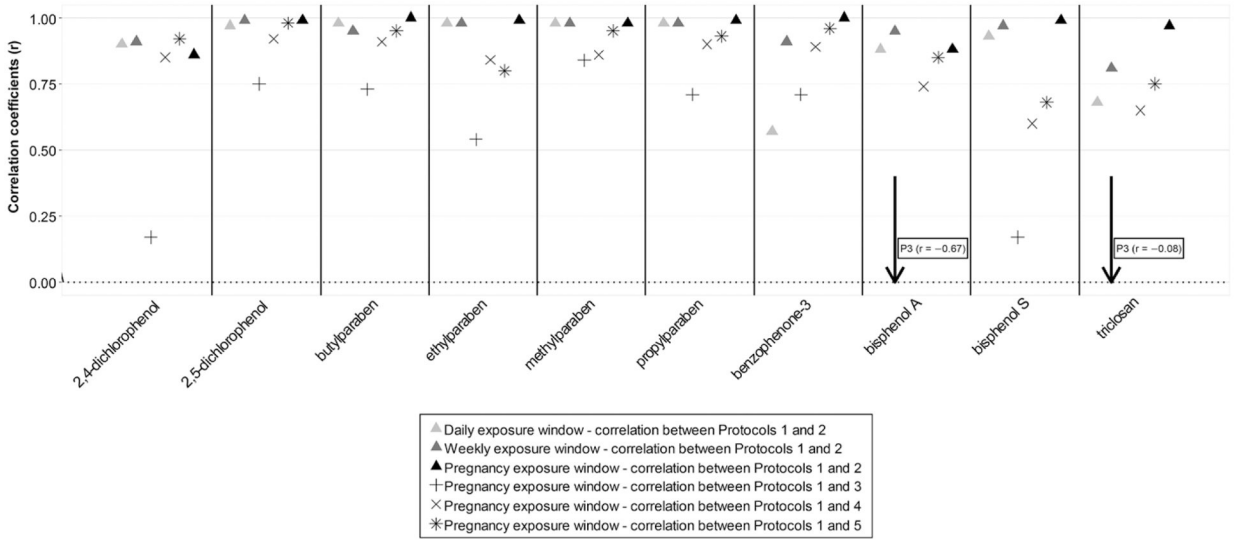

FIGURE 2.

All exposure windows: Pearson correlation coefficients (*r*) between protocols 1 (equal volumes of all spot urine voids were pooled within-subject) and 2 (equal volumes of three spot urine voids were pooled within-subject for daily pools, triangle marks) for all time windows: exposure estimates over a day (n = 8 women, N = 20 daily averages), a week (n = 8 women, N = 24 weekly averages), and the whole pregnancy (based on three measurement weeks, n = 16 women, N = 16 pregnancy averages). For the whole pregnancy, estimates from protocols 3 (+), 4 (×), and 5 (*) are also reported (eight women, eight pregnancy averages).

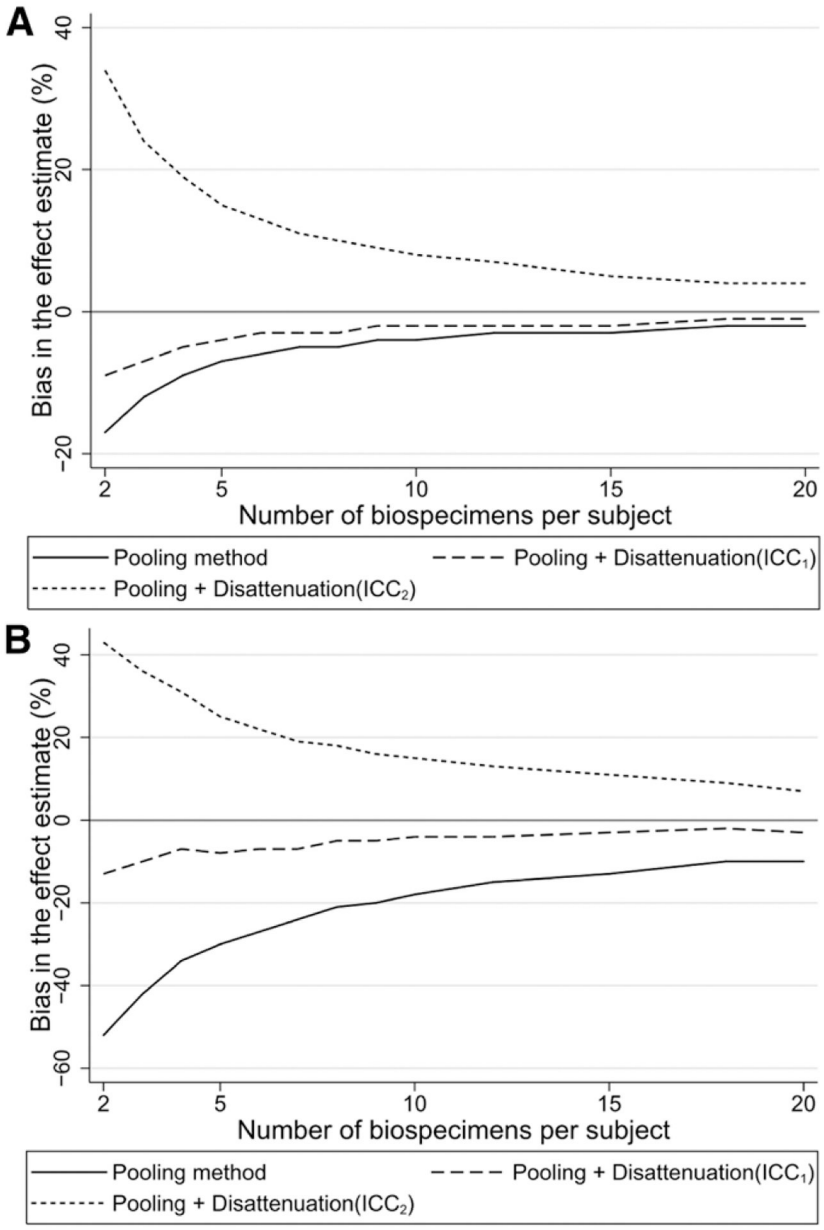

FIGURE 3.

Bias in the health effect estimate (in %) depending on the number of biospecimens pooled per subject to assess exposure (1,000 simulation runs with 3,000 subjects each; continuous health outcome, true effect β true = −100 g). A, Methylparaben (ICC_1 of 0.85 and ICC_2 of 0.45). B, Bisphenol A (ICC_1 of 0.38 and ICC_2 of 0.2).

TABLE 1.

Daily Exposure Window

Exposure Biomarker	LOD (μg/L)	Protocol 1 Daily Pools[a]				Protocol 2 Daily Pools[b]				Agreement Between Protocols 1 and 2				
			Percentiles (μg/L)				Percentiles (μg/L)							
		% >LOD	5th	50th	95th	% >LOD	5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P* [c]	ICC[d]
2,4-dichlorophenol	0.1	90	<LOD	0.30	3.35	100	0.20	0.25	2.85	0.90	0.61	0.22	0.37	0.12 (0.00, 0.28)
2,5-dichlorophenol	0.1	95	0.15	0.65	146	85	<LOD	0.45	110	0.97	0.90	0.54	0.06	0.11 (0.00, 0.27)
Butylparaben	0.1	85	<LOD	1.15	74.5	85	<LOD	0.85	94.5	0.98	0.97	0.77	0.76	0.10 (0.00, 0.25)
Ethylparaben	1.0	75	<LOD	4.70	53.6	80	<LOD	4.65	68.4	0.98	0.97	1.00	0.86	0.03 (0.00, 0.15)
Methylparaben	1.0	100	5.55	101.35	2,550	100	5.45	94.25	2,070	0.98	0.96	0.85	0.61	0.27 (0.05, 0.49)
Propylparaben	0.1	100	0.25	15.30	299	100	0.30	17.60	322	0.98	0.98	0.85	0.58	0.28 (0.05, 0.50)
Benzophenone-3	0.2	45	<LOD	<LOD	19.5	30	<LOD	<LOD	24.5	0.57	0.44	0.26	0.25	0.26 (0.04, 0.48)
Bisphenol A	0.1	100	0.65	1.55	6.65	100	0.65	2.00	6.95	0.88	0.84	0.62	0.06	0.21 (0.01, 0.41)
Bisphenol S	0.1	95	0.15	0.30	1.70	95	0.15	0.40	2.05	0.93	0.88	0.55	<0.001	0.50 (0.26, 0.73)
Triclosan	1.0	80	<LOD	2.55	45.8	50	<LOD	1.05	73.5	0.68	0.48	0.10	<0.001	0.30 (0.08, 0.53)
Urine dilution markers														
Creatinine (mg/dl)	NA	NA	41.1	96.5	154	NA	72.7	100	144	0.81	0.80	0.40	0.01	0.10 (0.00, 0.26)
Specific gravity (unitless)	NA	NA	1.013	1.018	1.030	NA	1.013	1.020	1.023	0.65	0.68	0.44	0.48	0.03 (0.00, 0.15)

Descriptive statistics of the biomarker concentrations for the 20 daily pools (eight women) and agreement between estimates from protocols 1 (pooling of all urine samples/day) and 2 (pooling of three urine samples/day). Biomarker concentrations were ln-transformed for the estimation of agreement. *r* indicates Pearson correlation coefficient, *ρ* indicates Spearman correlation coefficient, *K* indicates Kappa coefficient (based on biomarker concentration categorized into tertiles).

[a] All spot individual urine specimens of a day were pooled within-subject in equal volumes to obtain daily pools.

[b] Three individual spot urine specimens of a day were pooled within-subject in equal volumes to obtain daily pools.

[c] *P* value of Student *t* test comparing biomarker ln-transformed concentrations between daily pools from protocols 1 and 2.

[d] Within-day ICC, as reported in Vernet et al.[3]

LOD indicates limit of detection; NA, not applicable.

TABLE 2.

Weekly Exposure Window

Exposure Biomarker	LOD (μg/L)	Protocol 1 Weekly Pools[a]				Protocol 2 Weekly Pools[b]				Agreement Between Protocols 1 and 2				
			Percentiles (μg/L)				Percentiles (μg/L)							
		% >LOD	5th	50th	95th	% >LOD	5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P* [c]	ICC[d]
2,4-dichlorophenol	0.1	92	<LOD	0.30	2.20	100	0.20	0.30	2.10	0.91	0.80	0.50	0.49	0.91 (0.82, 1.00)
2,5-dichlorophenol	0.1	100	0.30	0.60	73.0	100	0.30	0.75	72.4	0.99	0.97	0.94	0.04	0.98 (0.95, 1.00)
Butylparaben	0.1	58	<LOD	0.25	24.0	63	<LOD	0.40	32.0	0.95	0.88	0.81	0.15	0.80 (0.61, 0.99)
Ethylparaben	1.0	79	<LOD	11.40	56.5	88	<LOD	9.20	56.8	0.98	0.96	0.88	0.48	0.85 (0.70, 1.00)
Methylparaben	1.0	100	3.90	44.40	1,670	100	5.45	51.1	1,120	0.98	0.95	0.75	0.11	0.84 (0.69, 1.00)
Propylparaben	0.1	96	0.20	4.80	174	100	0.30	6.05	123	0.98	0.95	1.00	0.01	0.90 (0.80, 1.00)
Benzophenone-3	0.2	38	<LOD	<LOD	28.5	42	<LOD	<LOD	36.7	0.91	0.89	0.76	0.37	0.73 (0.50, 0.96)
Bisphenol A	0.1	100	0.50	1.90	5.70	100	0.80	2.00	6.30	0.95	0.95	0.75	0.02	0.60 (0.30, 0.89)
Bisphenol S	0.1	92	<LOD	0.30	14.4	96	0.20	0.40	18.0	0.97	0.93	0.63	0.07	0.14 (0.00, 0.39)
Triclosan	1.0	79	<LOD	2.50	83.7	88	<LOD	2.35	96.1	0.81	0.83	0.44	0.93	0.89 (0.78, 1.00)
Urine dilution markers														
Creatinine (mg/dl)	NA	NA	49.6	84.1	142	NA	64.2	88.0	135	0.81	0.80	0.40	0.01	0.60 (0.30, 0.89)
Specific gravity (unitless)	NA	NA	1.009	1.016	1.023	NA	1.011	1.016	1.024	0.65	0.68	0.44	0.48	0.61 (0.32, 0.90)

Descriptive statistics of the biomarker concentrations for the 24 weekly pools (eight women) and agreement between estimates from protocols 1 (pooling of all urine samples/day) and 2 (pooling of three urine samples/day). Biomarker concentrations were ln-transformed for the estimation of agreement. *r* indicates Pearson correlation coefficient, *ρ* indicates Spearman correlation coefficient, *K* indicates Kappa coefficient (based on biomarker concentration categorized into tertiles).

[a] All individual spot urine specimens of a day were pooled within-subject in equal volume to obtain daily pools. Daily pools were pooled within-subject in equal volumes to create weekly pools.

[b] Three individual spot urine specimens of a day were pooled within-subject in equal volumes for daily pools. Daily pools were pooled within-subject in equal volumes to create weekly pools.

[c] *P* value of Student *t* test comparing biomarker ln-transformed concentrations from weekly pools from protocols 1 and 2.

[d] Within-week ICC, as reported in Vernet et al.[3]

LOD indicates limit of detection; NA, not applicable.

TABLE 3.

Pregnancy Exposure Window

Exposure Biomarker	Protocol	N	LOD (μg/L)	% >LOD	Percentiles (μg/L)			Agreement With Estimates From Protocol 1				ICC[b]
					5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P*[a]	
2,4-dichlorophenol	Pregnancy pool, protocol 1[c]	16	0.1	88	0.20	0.30	2.70	Ref				0.50 (0.08, 0.92)
	Pregnancy pool, protocol 2[d]	16		94	<LOD	0.30	2.40	0.86	0.88	0.89	0.43	
	Average of 8 random spot samples, protocol 5	8		NA	0.15	0.36	1.80	0.92	0.86	0.43	0.44	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	0.35	1.82	0.85	0.80	0.24	0.32	
	Single random spot sample, protocol 3	8		75	<LOD	0.40	0.50	0.17	0.15	−0.14	0.31	
2,5-dichlorophenol	Pregnancy pool, protocol 1[c]	16	0.1	100	0.40	0.55	117	Ref				0.85 (0.69, 1.00)
	Pregnancy pool, protocol 2[d]	16		100	0.30	1.00	103	0.99	0.97	1.00	0.52	
	Average of 8 random spot samples, protocol 5	8		NA	0.20	0.65	65.5	0.98	0.88	0.62	0.09	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	1.05	67.8	0.92	0.88	0.62	0.23	
	Single random spot sample, protocol 3	8		88	<LOD	0.70	13.6	0.75	0.39	0.05	0.19	
Butylparaben	Pregnancy pool, protocol 1[c]	16	0.1	75	<LOD	0.50	25.8	Ref				0.42 (0.00, 0.87)
	Pregnancy pool, protocol 2[d]	16		75	<LOD	0.50	32.0	1.00	0.99	0.81	0.75	
	Average of 8 random spot samples, protocol 5	8		NA	<LOD	0.20	23.4	0.95	0.92	1.00	0.04	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	0.30	17.6	0.91	0.86	0.62	0.52	
	Single random spot sample, protocol 3	8		38	<LOD	<LOD	40.5	0.73	0.44	0.24	0.42	
Ethylparaben	Pregnancy pool, protocol 1[c]	16	1.0	81	<LOD	9.35	55.2	Ref				0.40 (0.00, 0.85)
	Pregnancy pool, protocol 2[d]	16		88	<LOD	9.35	124	0.99	0.97	1.00	0.81	

Exposure Biomarker	Protocol	N	LOD (μg/L)	% >LOD	Percentiles (μg/L)			Agreement With Estimates From Protocol 1				
					5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P* [a]	ICC [b]
	Average of 8 random spot samples, protocol 5	8		NA	<LOD	1.79	55.1	0.80	0.57	0.43	0.00	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	1.94	40.1	0.84	0.57	0.43	0.00	
	Single random spot sample, protocol 3	8		50	<LOD	<LOD	154	0.54	0.40	0.05	0.03	
Methylparaben	Pregnancy pool, protocol 1 [c]	16	1.0	100	5.20	56.05	2,600	Ref				0.85 (0.68, 1.00)
	Pregnancy pool, protocol 2 [d]	16		100	12.30	55.35	2,950	0.98	0.99	0.81	0.08	
	Average of 8 random spot samples, protocol 5	8		NA	3.82	18.16	764	0.95	0.98	1.00	0.01	
	Average of 3 random spot samples, protocol 4	8		NA	2.57	11.36	331	0.86	1.00	1.00	0.03	
	Single random spot sample, protocol 3	8		88	<LOD	12.65	7,850	0.84	0.79	0.24	0.17	
Propylparaben	Pregnancy pool, protocol 1 [c]	16	0.1	100	0.20	7.30	289	Ref				0.70 (0.40, 1.00)
	Pregnancy pool, protocol 2 [d]	16		100	0.30	9.45	324	0.99	0.98	0.62	0.01	
	Average of 8 random spot samples, protocol 5	8		NA	0.18	2.66	98.7	0.93	0.92	0.62	0.05	
	Average of 3 random spot samples, protocol 4	8		NA	0.13	5.37	51.4	0.90	0.90	0.62	0.21	
	Single random spot sample, protocol 3	8		50	<LOD	0.70	1,180	0.71	0.65	0.62	0.10	
Benzophenone-3	Pregnancy pool, protocol 1 [c]	16	0.2	31	<LOD	<LOD	98.80	Ref				0.28 (0.00, 0.75)
	Pregnancy pool, protocol 2 [d]	16		31	<LOD	<LOD	173	1.00	1.00	1.00	0.29	
	Average of 8 random spot samples, protocol 5	8		NA	0.22	0.31	15.1	0.96	0.85	0.62	1.00	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	0.56	11.8	0.89	0.85	0.62	0.95	
	Single random spot sample, protocol 3	8		63	<LOD	1.50	9.00	0.71	0.70	0.43	0.85	
Bisphenol A	Pregnancy pool, protocol 1 [c]	16	0.1	100	0.70	2.45	4.50	Ref				0.38 (0.00, 0.83)

Epidemiology. Author manuscript; available in PMC 2022 April 07.

Exposure Biomarker	Protocol	N	LOD (μg/L)	% >LOD	Percentiles (μg/L)			Agreement With Estimates From Protocol 1				ICC[b]
					5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P* [a]	
	Pregnancy pool, protocol 2[d]	16		100	0.80	3.05	6.10	0.88	0.85	0.62	0.10	
	Average of 8 random spot samples, protocol 5	8		NA	0.56	1.71	2.57	0.85	0.79	0.62	0.01	
	Average of 3 random spot samples, protocol 4	8		NA	0.11	1.27	3.05	0.74	0.52	0.43	0.02	
	Single random spot sample, protocol 3	8		88	<LOD	1.20	10.7	−0.67	−0.49	−0.33	0.41	
Bisphenol S	Pregnancy pool, protocol 1[c]	16	0.1	94	0.20	0.45	7.30	Ref				0.33 (0.00, 0.80)
	Pregnancy pool, protocol 2[d]	16		94	<LOD	0.45	8.60	0.99	0.91	0.61	0.19	
	Average of 8 random spot samples, protocol 5	8		NA	0.16	0.31	4.72	0.68	0.62	0.24	0.10	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	0.27	3.28	0.60	0.59	0.24	0.09	
	Single random spot sample, protocol 3	8		50	<LOD	0.15	0.90	0.17	0.10	0.05	0.05	
Triclosan	Pregnancy pool, protocol 1[c]	16	1.0	100	1.70	5.00	248	Ref				0.11 (0.00, 0.58)
	Pregnancy pool, protocol 2[d]	16		100	<LOD	4.50	259	0.97	0.86	0.62	0.56	
	Average of 8 random spot samples, protocol 5	8		NA	<LOD	1.62	10.4	0.75	0.60	0.24	0.00	
	Average of 3 random spot samples, protocol 4	8		NA	<LOD	1.51	4.91	0.65	0.48	0.24	0.00	
	Single random spot sample, protocol 3	8		75	<LOD	1.20	2.00	−0.08	−0.19	−0.17	0.01	
Urine dilution markers												
Creatinine (mg/dl)	Pregnancy pool, protocol 1[c]	16	NA	NA	50.69	84.4	162	Ref				0.54 (0.14, 0.94)
	Pregnancy pool, protocol 2[d]	16		NA	62.99	87.5	159	0.95	0.93	0.62	0.16	
	Average of 8 random spot samples, protocol 5	8		NA	44.59	83.8	124	0.88	0.83	0.62	0.20	
	Average of 3 random spot samples, protocol 4	8		NA	18.03	70.5	163	0.87	0.74	−0.14	0.10	
	Single random spot sample, protocol 3	8		NA	15.34	88.6	229	−0.01	−0.05	0.24	0.57	

Exposure Biomarker	Protocol	N	LOD (μg/L)	% >LOD	Percentiles (μg/L)			Agreement With Estimates From Protocol 1				ICC[b]
					5th	50th	95th	Pearson (*r*)	Spearman (*ρ*)	Kappa (*K*)	*P* [a]	
Specific gravity (unitless)	Pregnancy pool, protocol 1[c]	16	NA	NA	1.009	1.016	1.025	Ref				0.69 (0.39, 1.00)
	Pregnancy pool, protocol 2[d]	16		NA	1.011	1.016	1.025	0.98	0.97	0.59	0.84	
	Average of 8 random spot samples, protocol 5	8		NA	1.010	1.017	1.022	0.91	0.83	0.62	0.77	
	Average of 3 random spot samples, protocol 4	8		NA	1.005	1.016	1.027	0.91	0.87	0.62	0.40	
	Single random spot sample, protocol 3	8		NA	1.005	1.013	1.024	0.67	0.66	0.41	0.19	

Descriptive statistics of the nontransformed biomarker concentrations for the entire pregnancy exposure window estimated by various exposure models and agreement between estimates from protocol 1 (pooling of all urine samples/day), and protocols 2 (pooling of 3 urine samples/day), 3 (on 1 random spot sample), 4 (mean of 3 random spot samples), and 5 (mean of 8 random spot samples). Biomarker concentrations were ln-transformed for the estimation of agreement. *r* indicates Pearson correlation coefficient, *ρ* indicates Spearman correlation coefficient, *K* indicates Kappa coefficient (based on biomarker concentration categorized into tertiles).

[a] *P* value of Student *t* test comparing biomarker ln-transformed concentrations between protocol 1 and the other protocols considered.

[b] Within-woman ICC based on 3 random spot samples, as reported in Vernet at al.[3]

[c] All individual spot urine specimens of a day were pooled within-subject in equal volumes to obtain daily pools. Daily pools were pooled within-subject in equal volumes to create weekly pools and the within-subject pregnancy pool was created by pooling equal volumes of weekly pools.

[d] Three individual spot urine specimens of a day were pooled within-subject in equal volumes for daily pools. Daily pools were pooled within-subject in equal volumes to create weekly pools and the within-subject pregnancy pool was created by pooling equal volumes of weekly pools.

LOD indicates limit of detection; NA, not applicable.