

Within-subject Pooling of Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies

Flavie Perrier,^{a,b} Lise Giorgis-Allemand,^{a,b} Rémy Slama,^{a,b} and Claire Philippat^{a,b,c,d}

Background: For chemicals with high within-subject temporal variability, assessing exposure biomarkers in a spot biospecimen poorly estimates average levels over long periods. The objective is to characterize the ability of within-subject pooling of biospecimens to reduce bias due to exposure misclassification when within-subject variability in biomarker concentrations is high.

Methods: We considered chemicals with intraclass correlation coefficients of 0.6 and 0.2. In a simulation study, we hypothesized that the chemical urinary concentrations averaged over a given time period were associated with a health outcome and estimated the bias of studies assessing exposure that collected 1 to 50 random biospecimens per subject. We assumed a classical type error. We studied associations using a within-subject pooling approach and two measurement error models (simulation extrapolation and regression calibration), the latter requiring the assay of more than one biospecimen per subject.

Results: For both continuous and binary outcomes, using one sample led to attenuation bias of 40% and 80% for compounds with intraclass correlation coefficients of 0.6 and 0.2, respectively. For a compound with an intraclass correlation coefficient of 0.6, the numbers of biospecimens required to limit bias to less than 10% were 6, 2, and 2 biospecimens with the pooling, simulation extrapolation, and regression calibration methods (these values were, respectively, 35, 8, and 2 for a compound with an intraclass correlation coefficient of 0.2). Compared with pooling, these methods did not improve power.

Conclusion: Within-subject pooling limits attenuation bias without increasing assay costs. Simulation extrapolation and regression

calibration further limit bias, compared with the pooling approach, but increase assay costs.

(*Epidemiology* 2016;27: 378–388)

Exposure assessment is a central issue in epidemiologic studies exploring the effects of environmental contaminants on human health. For chemicals with multiple exposure sources, biomarker measurements in biospecimens are often used to assess an exposure proxy.^{1,2} Such biomarker-based studies generally rely on few (often only one) biospecimens per subject. However, for chemicals with a short half-life, such as bisphenol A, phthalates, and pyrethroid pesticides, within-subject biomarker concentrations strongly vary over time.^{3–5} Consequently, a biomarker assay based on a single biospecimen is likely to imperfectly reflect the average exposure throughout long time periods (typically, a week to several years). In this setting, the biomarker concentration measured in a spot biospecimen varies around the true unmeasured value (corresponding to the average biomarker level during the toxicologically relevant exposure window) in a way such that the average of many replicate measurements is expected to approximate the true individual level. This corresponds to what is termed classical error.⁶ Classical error is expected to bias dose–response relationships toward the null in a predictable way^{7,8} and reduce statistical power.^{9–11}

Performing repeated exposure measurements on each study participant is generally a relevant option to reduce bias due to exposure misclassification in environmental health studies.^{7,8} This approach has so far little been used in epidemiologic studies on the health effect of chemicals with short half-lives.^{12,13} This might be due to the increased assay costs, and possibly to the assumption that increasing the number of study subjects is more efficient than increasing the number of biospecimens per subject (which is not always true¹⁴). One way to have some of the benefits of the reliance on repeated biospecimens per subject without increasing the assay costs, compared with the situation where only one biospecimen is collected, would be to collect and pool several biospecimens for each subject before assaying the chemical of interest (within-subject pooling). Pooling biospecimens from different subjects (between-subject pooling) has been applied since the 1940s to reduce the cost of identification of infectious cases in

Submitted 17 November 2014; accepted 1 February 2016.

From the ^aInserm, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, U1209 (IAB), Grenoble, France; ^bUniv. Grenoble-Alpes, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, IAB, Grenoble, France; ^cDepartment of Public Health Sciences, University of California, Davis, CA; and ^dMIND (Medical Investigations of Neurodevelopmental Disorders) Institute, University of California, Davis, CA.

This project is part of E-DOHaD project, funded by a Consolidator grant from the European Research Council (ERC Grant #StG2012-311765).

The authors report no conflict of interest.

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

Correspondence: Rémy Slama, Institut Albert Bonniot-Centre de Recherche INSERM-UJF U823, UJF Site Santé, BP 170, La Tronche, Grenoble 38042, Cedex 9, France. E-mail: remy.slama@ujf-grenoble.fr.

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

ISSN: 1044-3983/16/2703-0378

DOI: 10.1097/EDE.0000000000000460

large populations.^{15,16} More recently, in the context of case–control studies with expensive biomarker assays, between-subject pooling of biospecimens from random groups of cases and controls has been used to estimate individual effect of exposure while reducing assay costs.¹⁷ To the best of our knowledge, in spite of its simplicity, *within-subject* pooling of biospecimens has been very little used in biomarker-based environmental epidemiology. Notably, within-subject pooling has been proposed to limit the proportion of biospecimens with biomarker concentrations below the limit of detection,¹⁸ and as a strategy to estimate intraclass correlation coefficients (ICC).¹⁶

Our aim was therefore to assess the ability of within-subject pooling of biospecimens to reduce bias due to exposure misclassification occurring when biomarkers with strong temporal variability are considered. We compared the performance of this pooling approach to measurement error models, as approaches supposed to make optimal use of the collection of repeated biospecimens to adjust for measurement error.

METHODS

Overview of the Approach

We supposed that one is interested in estimating the health effects of exposure during a specific time-window to two distinct nonpersistent chemicals with different temporal variability in a biological matrix such as urine. We used different approaches to study the associations between the biomarker concentrations assayed with error in repeated biospecimens in each subject and continuous (e.g., child weight at the age of 3 years) and binary (e.g., being overweight) health outcomes: within-subject pooling of biospecimens before assaying the chemical (or *pooling method*, possibly followed by a posteriori disattenuation^{7,8}) and two measurement error models relying on the assay of the chemical in each biospecimen: regression calibration and simulation extrapolation (SIMEX). We simulated epidemiologic studies conducted in the general population and estimated bias and statistical power for the four approaches.

Simulation of Exposures

For each subject i ($i = 1, \dots, n$) a variable X_i , drawn at random from a normal distribution (mean, 0; standard deviation (SD) $\sigma_X = 1$), was assumed to represent the real unobserved average concentration (in $\mu\text{g}/\text{liter}$) of the considered chemical in the biospecimens collected over a toxicologically relevant exposure window, measured without error, after standardization and centering around 0.

For each subject, we generated 50 error-prone variables W_{ij} ($j = 1, \dots, 50$) corresponding to exposure biomarker concentrations assayed in spot biospecimens collected at random time points during the relevant exposure window. W_{ij} is an error-prone estimate of X_i . The within-subject error ($U_{ij, \text{within}}$, mean 0; SD, $\sigma_{U_{\text{within}}}$) affecting W_{ij} was supposed to be related to the individual and might be due to the variability in urinary concentrations arising from the toxicokinetics of the studied

compound, temporal variations in diluteness of the urine and temporal variability in exposure itself. $U_{ij, \text{within}}$ was assumed to correspond to random variations around the real value X_i , to be additive and of classical type. The variance of the error term $U_{ij, \text{within}}$ was computed as¹⁹

$$\sigma_{U_{\text{within}}}^2 = \sigma_X^2 \left(\frac{1}{\text{ICC}} - 1 \right), \quad (1)$$

where ICC is the intraclass correlation coefficient, corresponding to the expected correlation between biomarker concentrations in any pair of biospecimens W_{ij} and W_{ik} collected at different time points during the exposure window of interest. ICCs were set at 0.6 (moderate within-subject variability, chemical A) or 0.2 (high within-subject variability, chemical B). ICCs of about 0.6 have been typically reported for benzo-phenone-3, parabens, and a butylbenzylphthalate metabolite, while ICCs of 0.1–0.2 have been reported for bisphenol A and di-(2-ethylhexyl) phthalate (DEHP) metabolites.^{20,21}

In additional analyses, we also considered between-assay error ($U_{ij, \text{assay}}$) due to the biomarker measurement and arising from the technician and from analytical error (see eTables 1, 2; <http://links.lww.com/EDE/B27>).

In most simulations, we assumed the number of biospecimens to be the same for each subject (balanced design). We additionally specifically considered the situation of an unbalanced number of biospecimens per subject keeping constant the total number of subjects ($n = 3,000$) as well as the total number of biospecimens available in the population ($n = 9,000$).

Simulation of Health Outcomes

We simulated the continuous outcome Y_i as

$$Y_i = \alpha + \beta_1 X_i + \varepsilon_i, \quad (2)$$

where β_1 , the “real” effect of the chemical on Y , was assumed to be -100 g for each increase by 1 in X and ε was a normally distributed random error (mean = 0). The values of α (14,900 g) and of the standard deviation of ε (1,650 g) were based on the values observed for the weight at 3 years in a mother-child cohort (Eden cohort,²²). To study the impact of measurement error on the risk of type 1 error, we also conducted simulations assuming a lack of effect of exposure ($\beta_1 = 0$).

We also simulated the case of a binary health outcome (see eTables 3, 4; <http://links.lww.com/EDE/B27>).

Characterizing Bias in Studies Relying on Error-prone Estimates of Exposure

Within-subject Pooling

This method consisted of regressing the outcome Y over an exposure variable corresponding to the biomarker concentration measured in the pool of k ($k = 1, \dots, 50$) random (error-prone) biospecimens from each subject. We assumed that all individual specimens were pooled in equal volume, and errors that could arise from pooling volumes of individual

specimens that were not exactly equal were not considered. Associations between the concentration measured in the pool (average of $W_{i1} \dots W_{ik}$) and each health outcome were characterized by linear and logistic regression models for the continuous and binary health outcomes, respectively.

In the situations of unbalanced numbers of biospecimens pooled per participant, we used *weighted* linear and logistic regression models to assess associations between exposure and outcome. We used analytical weights depending on the number of biospecimens available for each participant; these weights were inversely proportional to the variance of the subject-specific exposure average (option *aweight* in Stata).

A Posteriori Disattenuation

In the balanced design, we additionally corrected the effect estimates from the pooling method^{7,8}:

$$\hat{\beta}_{\text{corr}} = \hat{\beta}_{\text{obs}} \cdot \left(\frac{k-1}{k} + \frac{1}{k \text{ICC}} \right), \quad (3)$$

where $\hat{\beta}_{\text{obs}}$ was the observed effect estimate with the pooling method, ICC the intraclass correlation coefficient, and k the number of biospecimens pooled per subject. The same correction was applied to the standard deviation.

Measurement Error Models

We considered two measurement error models (SIMEX and regression calibration^{11,23}), which make use of the assay of several biospecimens per subject to correct the dose–response function. For SIMEX, we used a quadratic model for the extrapolation step. We used bootstrap with 100 replications to estimate the variance of the estimated effects for both models. The models were implemented in Stata using the commands *rcal* for regression calibration²⁴ and *simex* for simulation extrapolation²⁵ (<http://www.stata.com/merror>; see eAppendix 1 for more details; <http://links.lww.com/EDE/B27>).

For each of the two chemicals considered, 1,000 studies were simulated. For each method, we quantified the average effect estimate ($\hat{\beta}_1$) and the statistical power, defined as the proportion of studies in which the P value of the parameter characterizing the association between the error-prone variables and the health outcome was below 0.05. Results are presented for a sample size of 3,000 subjects and with the default assumption of lack of between-assay error, unless otherwise specified.

Simulations and analyses were performed using STATA/SE, version 13 (StataCorp, College Station, TX). Our code for the continuous outcome is provided in the Supplemental Material (eAppendix 2, the code for the binary outcome is available upon request; <http://links.lww.com/EDE/B27>).

RESULTS

Distributions of X , W , and $U_{ij, \text{within}}$ and the simulated concentrations for three subjects are shown in eFigures 1 and 2 (<http://links.lww.com/EDE/B27>).

Bias Resulting from the Use of One Urine Sample to Assess Exposure

For chemical A (ICC of 0.6), the reliance on one urine sample to assess exposure led to an estimated effect of -60 g (average over 1,000 studies), corresponding to an attenuation by 40% compared with the real effect ($\beta_1 = -100$ g). Statistical power was 71% with a sample size of 3,000 subjects (Table 1). For chemical B (ICC of 0.2), when using one sample, the average effect estimate was -20 g (attenuation bias of 80%) and power was 32% (Table 2). Bias in the logarithm of the estimated odds-ratio associated with the exposure variable and power were similar for the binary outcome (eTables 3, 4; <http://links.lww.com/EDE/B27>).

For the continuous outcome, dividing the observed effect estimates by the ICC (a posteriori disattenuation^{7,8}) yielded corrected effect estimates of -100 g on average (i.e., no bias) for both chemicals. A posteriori disattenuation did not improve power (Tables 1, 2).

Increasing the Number of Biospecimens

Using the concentration assayed in the pool of several biospecimens per subject as a proxy of exposure decreased bias (Figure) and increased power, compared with the situation where one biospecimen was used per participant and no a posteriori disattenuation was applied. For both the continuous (Tables 1, 2) and binary (eTables 3, 4; <http://links.lww.com/EDE/B27>) outcomes, the number of biospecimens per subject required to limit the bias in the health effect estimate to 10% was 6 for chemical A (ICC of 0.6, Table 1, eTable 3; <http://links.lww.com/EDE/B27>) and 35 for chemical B (ICC of 0.2, Table 2, eTable 4; <http://links.lww.com/EDE/B27>). These values were reduced to 1 biospecimen when a posteriori disattenuation correction was applied.

Applying Measurement Error Models

For a given number of biospecimens, SIMEX and regression calibration drastically reduced bias compared with pooling without disattenuation (Figure). As an example, for an ICC of 0.6 and considering a continuous health outcome, using two biospecimens per subject led to effect estimates biased by 25%, 5%, and 0% with pooling, SIMEX, and regression calibration, respectively (Table 1); these values were respectively 66%, 45%, and 2% for a compound with an ICC of 0.2 (Table 2). Statistical power was similar (slightly higher) for SIMEX, regression calibration, and the pooling approach: for chemical B (ICC of 0.2) and considering a continuous health outcome, collecting two biospecimens led to a power of 50% for the pooling method, 52% for SIMEX, and 53% for regression calibration (study sample size of 3,000, Table 2).

Impact of Exposure Misclassification on Type 1 Error

When the real effect of exposure was assumed to be null and one biospecimen was used to assess exposure, the P value of the parameter associated with exposure was below 0.05 for

TABLE 1. Effect Estimates and Statistical Power of Studies Aiming at Characterizing the Association Between Biomarker-based Exposure to Chemical A (Intraclass Correlation Coefficient, 0.6) and a Continuous Health Outcome, According to the Number of Biospecimens Collected Per Subject and the Approach Used to Limit the Effect of Exposure Misclassification (1,000 Simulations of Studies with 3,000 Subjects Each; Real Effect $\beta_1 = -100$ g, Assuming Lack of Between-assay Error)

Number of Biospecimens Per Subject	Within-subject Pooling						Within-subject Pooling + A Posteriori Dissatenuation						SIMEX						Regression Calibration						
	Real Effect ^a	Effect Estimate ^a (95% CI) ^b		Power ^c	Bias (%) ^d	Bias (%) ^d	Effect Estimate ^a (95% CI) ^b	Power ^c	Bias (%) ^d	Bias (%) ^d	Effect Estimate ^a (95% CI) ^b	Power ^c	Bias (%) ^d	Bias (%) ^d	Effect Estimate ^a (95% CI) ^b	Power ^c	Bias (%) ^d	Bias (%) ^d	Effect Estimate ^a (95% CI) ^b	Power ^c	Bias (%) ^d	Bias (%) ^d			
1	-100	-60 ^e	(-105, -15)	0.71	40	-100 ^e	(-176, -24)	0.71	0	-100 ^e	(-176, -24)	0.71	0	-100 ^e	(-176, -24)	0.71	0	-100 ^e	(-176, -24)	0.71	0	-100 ^e	(-176, -24)	0.71	0
2	-100	-75	(-127, -22)	0.82	25	-100	(-169, -30)	0.82	0	-100	(-169, -30)	0.82	0	-100	(-169, -30)	0.82	0	-100	(-169, -30)	0.82	0	-100	(-169, -30)	0.82	0
3	-100	-82	(-139, -30)	0.85	18	-100	(-170, -36)	0.85	0	-100	(-170, -36)	0.85	0	-100	(-170, -36)	0.85	0	-100	(-170, -36)	0.85	0	-100	(-170, -36)	0.85	0
4	-102	-88	(-143, -33)	0.88	14	-102	(-167, -38)	0.88	0	-102	(-167, -38)	0.88	0	-102	(-167, -38)	0.88	0	-102	(-167, -38)	0.88	0	-102	(-167, -38)	0.88	0
5	-100	-89	(-145, -31)	0.88	11	-100	(-164, -35)	0.88	0	-100	(-164, -35)	0.88	0	-100	(-164, -35)	0.88	0	-100	(-164, -35)	0.88	0	-100	(-164, -35)	0.88	0
6	-99	-89	(-143, -35)	0.87	10	-99	(-158, -38)	0.87	0	-99	(-158, -38)	0.87	0	-99	(-158, -38)	0.87	0	-99	(-158, -38)	0.87	0	-99	(-158, -38)	0.87	0
7	-99	-91	(-149, -33)	0.89	8	-99	(-163, -36)	0.89	0	-99	(-163, -36)	0.89	0	-99	(-163, -36)	0.89	0	-99	(-163, -36)	0.89	0	-99	(-163, -36)	0.89	0
8	-98	-91	(-147, -36)	0.87	8	-98	(-159, -38)	0.87	0	-98	(-159, -38)	0.87	0	-98	(-159, -38)	0.87	0	-98	(-159, -38)	0.87	0	-98	(-159, -38)	0.87	0
9	-101	-95	(-151, -37)	0.90	7	-102	(-162, -39)	0.90	1	-102	(-162, -39)	0.90	1	-101	(-162, -39)	0.91	0	-101	(-162, -39)	0.91	0	-102	(-162, -40)	0.92	0
10	-99	-93	(-150, -38)	0.90	6	-100	(-160, -40)	0.90	1	-100	(-160, -40)	0.90	1	-100	(-159, -38)	0.91	0	-100	(-160, -40)	0.91	0	-100	(-160, -40)	0.91	0
12	-100	-95	(-153, -35)	0.89	5	-100	(-162, -37)	0.89	0	-100	(-162, -37)	0.89	0	-100	(-161, -39)	0.91	0	-100	(-162, -37)	0.91	0	-100	(-162, -37)	0.91	0
15	-99	-95	(-151, -39)	0.91	4	-99	(-158, -40)	0.91	0	-99	(-158, -40)	0.91	0	-100	(-157, -39)	0.92	0	-100	(-158, -40)	0.92	0	-100	(-158, -40)	0.92	0
18	-100	-96	(-156, -40)	0.91	4	-100	(-162, -42)	0.91	0	-100	(-162, -42)	0.91	0	-100	(-162, -41)	0.92	0	-100	(-162, -42)	0.92	0	-100	(-162, -42)	0.92	0
20	-99	-96	(-155, -38)	0.90	3	-99	(-161, -40)	0.90	0	-99	(-161, -40)	0.90	0	-99	(-161, -38)	0.91	0	-99	(-161, -40)	0.91	0	-99	(-161, -40)	0.92	0
25	-99	-97	(-158, -41)	0.91	3	-99	(-162, -42)	0.91	0	-99	(-162, -42)	0.91	0	-99	(-161, -42)	0.93	0	-99	(-162, -42)	0.93	0	-99	(-162, -42)	0.93	0
30	-101	-99	(-155, -40)	0.91	2	-101	(-158, -40)	0.91	0	-101	(-158, -40)	0.91	0	-101	(-157, -39)	0.92	0	-101	(-158, -40)	0.92	0	-101	(-158, -40)	0.93	0
35	-102	-100	(-155, -41)	0.91	2	-102	(-158, -42)	0.91	0	-102	(-158, -42)	0.91	0	-102	(-159, -43)	0.93	0	-102	(-158, -42)	0.93	0	-102	(-158, -42)	0.93	0
40	-101	-99	(-162, -40)	0.90	2	-101	(-165, -41)	0.90	0	-101	(-165, -41)	0.90	0	-101	(-165, -42)	0.92	0	-101	(-165, -41)	0.92	0	-101	(-165, -41)	0.92	0
50	-99	-97	(-154, -40)	0.90	2	-98	(-156, -40)	0.90	1	-98	(-156, -40)	0.90	1	-98	(-158, -42)	0.92	0	-98	(-156, -41)	0.92	0	-98	(-156, -41)	0.92	0

SIMEX and regression calibration cannot be used if only one biospecimen per subject is available.

^aMean of 1,000 effect estimates.

^bEmpirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1,000 simulation runs.

^cStatistical power, estimated as the proportion of studies in which the *P* value of the parameter characterizing the association between the error-prone exposure variable (W_i) and the continuous outcome was below 0.05.

^dDifference between the real effect and the effect estimate, divided by the real effect.

^eCorresponds to a situation without pooling.

TABLE 2. Effect Estimates and Statistical Power of the Simulated Studies Aiming at Characterizing the Association Between Biomarker-based Exposure to Chemical B (Intraclass Correlation Coefficient, 0.2) and a Continuous Health Outcome, According to the Number of Biospecimens Collected Per Subject and the Approach Used to Limit Exposure Misclassification (1,000 Simulations of Studies with 3,000 Subjects Each; Real Effect $\beta_1 = -100$ g, Assuming Lack of Between-assay Error)

Number of Biospecimens Per Subject	Within-subject Pooling			Within-subject Pooling + A Posteriori Dissatenuation			SIMEX			Regression Calibration			
	Real Effect ^a	Effect Estimate ^a (95% CI ^b)	Power ^c	Bias (%) ^d	Effect Estimate ^a (95% CI ^b)	Power ^c	Bias (%) ^d	Effect Estimate ^a (95% CI ^b)	Power ^c	Bias (%) ^d	Effect Estimate ^a (95% CI ^b)	Power ^c	Bias (%) ^d
1	-100	-20 ^e (-45, 8)	0.32	80	-100 ^e (-223, 38)	0.32	0	-54 (-110, 3)	0.52	45	-101 (-202, 6)	0.53	2
2	-100	-34 (-67, 2)	0.50	66	-101 (-201, 5)	0.50	1	-66 (-129, -3)	0.61	34	-100 (-196, -7)	0.62	1
3	-100	-43 (-83, -3)	0.58	58	-99 (-194, -7)	0.58	1	-77 (-139, -13)	0.70	24	-103 (-187, -17)	0.72	1
4	-102	-52 (-92, -9)	0.68	49	-103 (-184, -17)	0.68	1	-82 (-151, -19)	0.72	18	-102 (-188, -23)	0.74	2
5	-100	-56 (-105, -13)	0.71	43	-101 (-188, -23)	0.71	1	-84 (-151, -23)	0.74	15	-100 (-178, -27)	0.74	0
6	-99	-60 (-108, -16)	0.70	40	-99 (-180, -27)	0.70	0	-88 (-155, -19)	0.79	11	-100 (-178, -23)	0.80	1
7	-99	-64 (-113, -15)	0.77	35	-100 (-177, -23)	0.77	1	-88 (-152, -21)	0.78	10	-98 (-170, -25)	0.78	0
8	-98	-66 (-112, -17)	0.74	33	-98 (-168, -25)	0.74	0	-93 (-159, -27)	0.84	8	-102 (-174, -28)	0.84	0
9	-101	-71 (-121, -20)	0.81	30	-102 (-174, -28)	0.81	1	-93 (-154, -29)	0.82	6	-100 (-167, -33)	0.83	1
10	-99	-72 (-119, -24)	0.80	28	-100 (-166, -33)	0.80	1	-95 (-163, -25)	0.83	5	-100 (-172, -28)	0.84	0
12	-100	-75 (-129, -22)	0.82	25	-100 (-172, -29)	0.82	0	-97 (-158, -33)	0.88	3	-100 (-164, -37)	0.88	1
15	-99	-79 (-130, -29)	0.86	21	-100 (-164, -37)	0.86	1	-97 (-161, -34)	0.87	3	-100 (-165, -37)	0.87	0
18	-100	-81 (-135, -30)	0.85	18	-99 (-165, -37)	0.85	1	-97 (-162, -31)	0.88	2	-99 (-166, -35)	0.88	0
20	-99	-83 (-139, -29)	0.85	16	-99 (-167, -35)	0.85	0	-98 (-161, -36)	0.89	1	-99 (-162, -36)	0.91	0
25	-99	-85 (-139, -31)	0.88	14	-99 (-161, -36)	0.88	0	-100 (-161, -35)	0.89	1	-101 (-161, -35)	0.89	0
30	-101	-89 (-143, -31)	0.88	12	-101 (-162, -35)	0.88	0	-101 (-161, -40)	0.91	0	-102 (-161, -43)	0.91	0
35	-102	-91 (-145, -39)	0.89	10	-102 (-161, -43)	0.89	0	-100 (-168, -38)	0.90	1	-101 (-168, -38)	0.90	0
40	-101	-92 (-153, -35)	0.88	9	-101 (-168, -38)	0.88	0	-98 (-156, -36)	0.90	1	-98 (-157, -37)	0.90	0
50	-99	-91 (-146, -34)	0.88	8	-98 (-157, -37)	0.88	1						

SIMEX and regression calibration cannot be used if only one biospecimen per subject is available.

^aMean of 1,000 effect estimates.

^bEmpirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1,000 simulation runs.

^cStatistical power, estimated as the proportion of studies in which the *P* value of the parameter characterizing the association between the error-prone variables (W_j) and the continuous outcome was below 0.05.

^dDifference between the real effect and the effect estimate, divided by the real effect.

^eCorresponds to a situation without pooling.

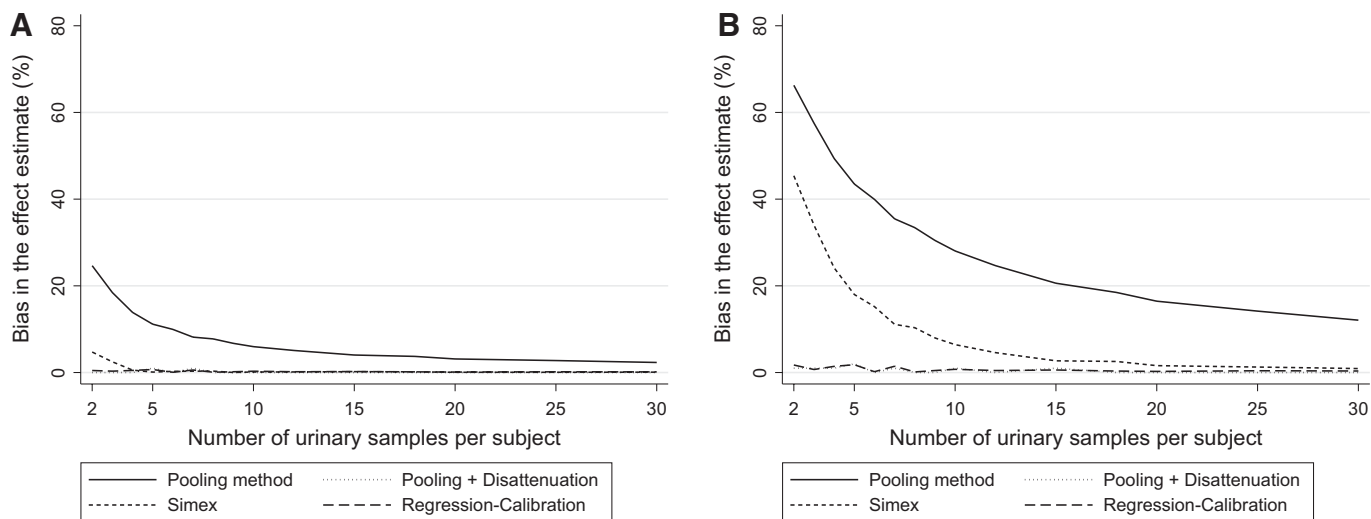


FIGURE. Bias in the health effect estimate (%) according to the number of biospecimens collected per subject to assess exposure (1,000 simulations of studies with 3,000 subjects each; continuous health outcome, real effect $\beta_1 = -100$ g, (assuming lack of between-assay error, $\sigma_{U_{\text{assay}}}^2 = 0$). A, Compound A (ICC of 0.6). B, Compound B (ICC of 0.2).

5% of the simulated datasets, whatever the ICC considered (eTables 5, 6; <http://links.lww.com/EDE/B27>), suggesting that under the hypotheses of our simulations, using one biospecimen to assess exposure did not increase the risk of type I error. For a given number of biospecimens, the risk of type I error was slightly higher with SIMEX and regression calibration (range: 5%–8%) compared with the pooling approach (range: 4%–6%, eTables 5, 6; <http://links.lww.com/EDE/B27>).

Unbalanced Number of Biospecimens Between Subjects

For chemical A (ICC of 0.6), and for a given total number of biospecimens collected (9,000 samples among 3,000 subjects), bias in the effect estimate was slightly higher when the design was unbalanced (different numbers of biospecimens per subject, bias 20%–22%) than when the design was balanced (bias of 18%, Table 3). For chemical B (ICC of 0.2), the effect estimate tended to be less biased with unbalanced (47%–52%) compared with balanced (58%) designs (Table 3). When measurement error models were applied, bias was always lower with a balanced design than with unbalanced designs (Table 3).

Impact of Study Sample Size

Overall, for a given number of biospecimens per participant, varying study sample size did not affect bias but impacted statistical power (Table 4). For both ICCs, we observed higher bias and higher statistical power in studies relying on one biospecimen per subject compared with studies with two pooled biospecimens and half the sample size (Table 4). For compounds with an ICC of 0.2, the loss of power when recruiting half as many subjects with twice as many biospecimens was sometimes limited: power was 37% for a study of 4,000 subjects with one biospecimen per subject, and 33% in a study of

2,000 subjects in which two biospecimens were pooled by subject; bias was lower (67%) in the latter study with 2,000 subjects than in the former one with 4,000 subjects (80%, Table 4).

Impact of Between-assay Error

For a given number of biospecimens, with the pooling method, bias increased with increasing between-assay error (eTable 1; <http://links.lww.com/EDE/B27>). When several biospecimens were available per subject, performance of a posteriori disattenuation was reduced with increasing between-assay error. For a compound with an ICC of 0.6, when five biospecimens were available per subject, a posteriori disattenuation led to effect estimates biased by 1%, 8%, and 31% for between-assay variances of 0.01, 0.1, and 0.5, respectively. With regression calibration, the between-assay variance did not affect bias while with SIMEX, bias increased with increasing between-assay error (eTable 1; <http://links.lww.com/EDE/B27>). This increase in bias is likely to result from the effect of the ICC decrease (from 0.60 to 0.46 and from 0.20 to 0.18) that resulted from the increase in the between-assay error, rather than result from a direct effect of the between assay error. Indeed, in simulations in which ICCs were kept constant whatever the value of the between-assay error, with SIMEX, bias was not affected by increases in the between-assay variance (eTable 2; <http://links.lww.com/EDE/B27>).

In situations of substantial between-assay error ($\sigma_{U_{\text{assay}}}^2$, 0.1 and 0.5), power was lower with pooling compared with SIMEX and regression calibration (eTables 1, 2; <http://links.lww.com/EDE/B27>).

DISCUSSION

Our simulation study confirmed that relying on one biospecimen to assess exposure to short half-life chemicals

TABLE 3. Unbalanced Designs: Effect Estimates and Statistical Power When the Numbers of Biospecimens Available Differed Between Subjects (Continuous Outcome, 1,000 Simulations for Each Design; Real Effect $\beta_1 = -100$, Assuming a Lack of Between-assay Error)

Design	True Effect Estimate ^b	Within-subject Pooling ^a			SIMEX			Regression Calibration		
		Effect Estimate ^b	Power ^c	Bias (%) ^d	Effect Estimate ^b	Power ^c	Bias (%) ^d	Effect Estimate ^b	Power ^c	Bias (%) ^d
Chemical A, intraclass correlation coefficient of 0.6										
a	-102	-81	0.81	21	-95	0.84	7	-101	0.85	1
b	-101	-79	0.83	22	-97	0.85	4	-102	0.86	1
c	-99	-79	0.82	20	-96	0.85	3	-102	0.85	3
d	-100	-80	0.78	20	-93	0.81	7	-101	0.82	1
e	-100	-82	0.85	18	-98	0.86	2	-100	0.86	0
Chemical B, intraclass correlation coefficient of 0.2										
a	-99	-48	0.62	51	-61	0.59	38	-105	0.64	6
b	-102	-55	0.69	47	-65	0.62	37	-96	0.71	6
c	-100	-50	0.63	50	-53	0.53	47	-87	0.65	13
d	-100	-48	0.62	52	-52	0.50	48	-107	0.59	7
e	-100	-43	0.58	58	-66	0.61	34	-100	0.62	1

Design a: 600 subjects with one sample, 1,500 with two samples, and 900 with six samples; design b: 600 subjects with one sample, 1,800 with three samples, and 600 with five samples; design c: 900 subjects with one sample, 1,200 with three samples, and 900 with five samples; design d: 1,200 subjects with one sample, 1,200 with two samples, and 600 with nine samples; design e: 3,000 subjects with three biospecimens each (balanced design).

^aWeighted linear regression models were used to assess associations between exposure and outcome.

^bMean of 1,000 effect estimates.

^cStatistical power, proportion of studies in which the *P* value of the parameter characterizing the association between the error-prone variables (W_{ij}) and the continuous outcome was below 0.05.

^dDifference between the real effect and the effect estimate, divided by the real effect.

yields biased dose–response functions. Increasing the number of biospecimens per subject and pooling them was efficient in decreasing bias and increasing statistical power without affecting assay costs, compared with the usual approach in which only one biospecimen per subject is collected. When the number of biospecimens pooled is identical for each subject, bias due to classical-type error could be further reduced by applying an a posteriori correction (disattenuation) to the effect estimate obtained with the pooling method.^{7,8}

Bias and Power in Studies Relying on One Biospecimen

The attenuation in the effect estimates observed in our study has been previously described in situations of measurement error of classical type.^{7,8,10,11} When one biospecimen per subject was used to assess exposure, our results were in line with the property that the ICC corresponds to the multiplicative attenuation factor in the parameter of the regression model.^{7,8} Consequently, if the ICC is known, dividing the observed effect estimate by the ICC (a posteriori disattenuation) provided unbiased estimate of the real effect. Such a correction did not improve power. For compounds with low ICCs, such as bisphenol A, DEHP metabolites and some pesticides (ICC below 0.3),^{21,26,27} the bias in studies relying on a single biospecimen per subject appears very large (more than 70% attenuation), making such studies of limited informative value regarding the dose–response function whatever the number of subjects. Bias remains large with an ICC of 0.6.

The decrease in power observed with decreasing ICC was coherent with the fact that the bias in the health effect estimates corresponded to attenuation.¹¹ As discussed elsewhere,²⁸ when several chemicals with different ICCs are simultaneously considered,^{29,30} interpretation of results should consider these ICCs, in particular when no association is highlighted. Without bias correction, such studies cannot easily be used to identify which exposure, among those tested, is most harmful, because the amplitude of the bias is likely to differ between exposures.

Increasing the Number of Biospecimens Per Subject

Collecting repeated biospecimens per subject is an efficient method to decrease bias. The number of biospecimens that has to be collected to restrict attenuation of the health effect estimate can be estimated from the formula from Brunekreef et al.,⁸ and Rappaport et al.⁷ (Eq. 3), which predicts that six and 36 biospecimens per participants would be needed to limit the attenuation factor to 0.9 for chemicals with ICCs of 0.6 and 0.2, respectively. This was in line with our estimations based on a simulation approach (six and more than 35 biospecimens, respectively). Most epidemiologic studies relying on biomarkers to assess exposure to short half-life compounds relied on a small number of biospecimens per subject.^{13,22,31} A possible explanation lies in the increased cost incurred by the assessment of biomarkers in several biospecimens per subject. This is not a justification to refrain from collecting several biospecimens per

TABLE 4. Effect Estimates and Statistical Power of the Simulated Studies Aiming at Characterizing the Association Between a Biomarker-based Exposure and a Continuous Health Outcome According to the Number of Biospecimens Collected Per Subject, the Population Size and the Method Used to Study the Associations (1,000 Simulations for Each Population Size; Real Effect $\beta_1 = -100$ g, Assuming a Lack of Between-assay Error)

Sample Size	Number of Specimens Per Subject	Real Effect ^a	Chemical A, ICC of 0.6						Chemical B, ICC of 0.2																	
			Within-subject Pooling		Within-subject Pooling + Disatenuation		Regression Calibration		Within-subject Pooling		Within-subject Pooling + Disatenuation		Regression Calibration													
			Effect Estimate ^a	Power ^b (%)	Bias (%) ^c	Effect Estimate ^a	Power ^b (%)	Bias (%) ^c	Effect Estimate ^a	Power ^b (%)	Bias (%) ^c	Effect Estimate ^a	Power ^b (%)	Bias (%) ^c	Effect Estimate ^a	Power ^b (%)	Bias (%) ^c									
500	1	-97	-60 ^d	0.18	38	-100 ^d	0.18	3	-96	0.25	6	-102	0.26	1	-34	0.10	79	-104 ^d	0.10	7	-104 ^d	0.10	7			
	2	-101	-76	0.22	25	-102	0.22	0	-96	0.25	6	-102	0.26	1	-34	0.12	66	-102	0.12	1	-55	0.14	46	-108	0.10	6
	5	-99	-86	0.24	13	-97	0.24	2	-95	0.27	4	-97	0.28	2	-53	0.16	47	-95	0.16	4	-75	0.18	24	-95	0.19	4
	10	-101	-95	0.27	6	-101	0.27	0	-101	0.31	1	-101	0.31	0	-72	0.22	29	-101	0.22	0	-93	0.24	8	-101	0.25	0
	50	-101	-100	0.29	1	-102	0.29	0	-102	0.34	0	-102	0.34	0	-94	0.27	7	-102	0.27	0	-101	0.30	0	-102	0.30	0
1,000	1	-100	-62 ^d	0.32	38	-103 ^d	0.32	3	-96	0.44	4	-101	0.44	1	-34	0.13	79	-107 ^d	0.13	6	-56	0.24	44	-105	0.22	5
	2	-100	-76	0.39	24	-101	0.39	1	-96	0.44	4	-101	0.44	1	-34	0.21	66	-103	0.21	3	-81	0.31	19	-101	0.32	1
	5	-100	-88	0.43	12	-100	0.43	0	-99	0.47	1	-100	0.47	0	-56	0.29	44	-100	0.29	1	-92	0.40	8	-99	0.42	1
	10	-101	-94	0.45	7	-100	0.45	1	-100	0.49	1	-100	0.51	0	-71	0.36	29	-99	0.37	1	-100	0.50	1	-101	0.50	0
	50	-101	-99	0.47	1	-101	0.47	0	-101	0.53	0	-101	0.52	0	-93	0.45	7	-101	0.46	0	-100	0.50	1	-101	0.50	0
2,000	1	-98	-60 ^d	0.56	39	-99 ^d	0.56	1	-95	0.69	6	-100	0.70	1	-33	0.22	80	-101 ^d	0.22	2	-53	0.39	47	-100	0.37	1
	2	-100	-75	0.66	25	-100	0.66	1	-95	0.69	6	-100	0.70	1	-33	0.33	67	-99	0.34	1	-81	0.58	19	-101	0.59	1
	5	-100	-89	0.72	11	-100	0.73	0	-99	0.75	0	-100	0.76	0	-56	0.54	44	-101	0.54	1	-94	0.67	6	-101	0.67	1
	10	-100	-94	0.75	6	-101	0.75	0	-101	0.78	0	-101	0.79	0	-72	0.63	28	-101	0.63	1	-99	0.78	0	-99	0.78	0
	50	-99	-98	0.77	1	-99	0.77	0	-99	0.81	0	-99	0.81	0	-92	0.74	7	-99	0.74	0	-99	0.78	0	-99	0.78	0
3,000	1	-100	-60 ^d	0.71	40	-100 ^d	0.71	0	-95	0.85	5	-100	0.86	0	-34	0.32	80	-100 ^d	0.32	0	-54	0.52	45	-101	0.53	2
	2	-100	-75	0.82	25	-100	0.82	0	-95	0.85	5	-100	0.86	0	-34	0.50	66	-101	0.50	1	-82	0.72	18	-102	0.74	2
	5	-100	-89	0.88	11	-100	0.88	0	-100	0.89	0	-101	0.91	1	-56	0.71	43	-101	0.71	1	-93	0.82	6	-100	0.83	1
	10	-99	-93	0.90	6	-100	0.90	1	-100	0.91	0	-100	0.91	0	-72	0.80	28	-100	0.80	1	-98	0.90	1	-98	0.90	0
	50	-99	-97	0.90	2	-98	0.90	1	-98	0.92	0	-98	0.92	0	-91	0.88	8	-98	0.88	1	-98	0.90	1	-98	0.90	0
4,000	1	-98	-58 ^d	0.82	40	-97 ^d	0.82	1	-94	0.93	6	-100	0.93	0	-33	0.37	80	-96 ^d	0.37	2	-53	0.63	47	-99	0.63	1
	2	-100	-75	0.92	25	-100	0.92	0	-94	0.93	6	-100	0.93	0	-33	0.59	67	-99	0.59	1	-82	0.85	19	-102	0.86	1
	5	-100	-89	0.96	11	-101	0.96	0	-100	0.97	0	-101	0.97	0	-56	0.83	44	-101	0.83	1	-93	0.91	7	-101	0.93	1
	10	-100	-94	0.96	6	-100	0.96	0	-100	0.97	0	-100	0.97	0	-72	0.91	28	-100	0.91	1	-93	0.91	7	-101	0.93	1
	50	-100	-99	0.97	1	-100	0.97	0	-100	0.98	0	-100	0.98	0	-93	0.96	7	-100	0.96	0	-100	0.96	0	-100	0.96	0

^aMean of 1,000 effect estimates.
^bStatistical power, proportion of studies in which the *P* value of the parameter characterizing the association between the error-prone variables (w_{ij}) and the continuous outcome was below 0.05.
^cDifference between the real effect and the effect estimates, divided by the real effect.
^dCorresponds to a situation without pooling.

subject. Indeed, the within-subject pooling approach described here allows to decrease bias and increase power without increasing assay costs, compared with the situation when one spot biospecimen is collected. When the number of biospecimens pooled is identical for each subject, a posteriori disattenuation^{7,8} can be used to further reduce attenuation bias. If the number of samples collected varies between subjects (unbalanced design), then the formula to disattenuate effect estimates a posteriori cannot rigorously be used anymore, but, as we have illustrated, the within-subject pooling approach still applies.

When no a posteriori correction of the effect estimate was performed, for a given number of biospecimens collected, studies relying on one biospecimen were more biased but had higher power than studies with two pooled biospecimens and half the sample size. For a compound with an ICC of 0.2, the gain in power was sometimes limited. For such compounds, compared with studies with a single biospecimen, studies with two or more samples per subject and possibly fewer subjects are an alternative worth considering. These would moreover have a lower assay cost.

Limitations of the Within-subject Pooling Approach

For chemicals with high intraindividual variability (ICC of 0.2), without a posteriori disattenuation, a large number of biospecimens (above 30) was required for the pooling approach to limit bias to 10%. Collecting more than 30 samples per subject outside of a clinical setting is cumbersome for the subjects (possibly limiting participation rate and inducing selection bias) and in terms of collection, storage and processing of the biological samples, which might impact the total cost of the cohort.

Caution is required in attempting to correct biomarker levels for creatinine concentration when pooling is done. The mathematical average of the creatinine-standardized biomarker concentrations (biomarker concentration/creatinine concentration) assessed in two biospecimens will indeed generally differ from the creatinine-standardized concentration assessed in the pool made of these two biospecimens; this feature might limit the efficiency of creatinine correction through standardization in the pooling approach. Other approaches could be considered to take into account urine dilution.³²

If the toxicologically relevant exposure window is unknown, pooling biospecimens over long exposure windows should be considered with caution, as pooling biospecimens collected during the relevant exposure window with biospecimens collected in another exposure window may increase exposure misclassification instead of decreasing it, in particular if exposure varied between the two windows considered.

Use of Measurement Error Models to Decrease Bias

Compared with the pooling approach used without a posteriori disattenuation, for a given number of biospecimens per subject, SIMEX and regression calibration strongly reduced bias. In our simulations, for chemicals with moderate

intraindividual variability (ICC of 0.6), both measurement error models behaved similarly and two biospecimens were enough to limit bias to 10%. For chemicals with high variability (ICC of 0.2), SIMEX appeared to be less able than regression calibration to reduce bias. In line with our results, studies relying on linear and logistic models overall observed more biased effect estimates with SIMEX than with regression calibration in the case of high intraindividual variability in exposure.^{33,34}

The slight gain in power observed with SIMEX and regression calibration compared with pooling is likely to result from the fact that, under the assumptions of our simulation, SIMEX and regression calibration led to inflated type I error, compared with pooling. The fact that, compared with pooling, measurement error models did not strongly improve power while reducing bias, is a manifestation of the *bias versus variance* tradeoff¹¹: measurement error models led to effect estimates that were less attenuated (further away from the null), but with larger variances than the effect estimates from the pooling approach.

SIMEX and regression calibration corrected for both the between-assay error and the error related to the individual, while the within-subject pooling method was inefficient in correcting for the between-assay error. An explanation is that, in contrast to the individual error, which tends to cancel out when several biospecimens are pooled, the between-assay error occurs after the pooling is done and remains the same whatever the number of biospecimens pooled. Conversely, SIMEX and regression calibration methods identified the total error (including the intraindividual and between-assay error) through the repeated assays and corrected for it.

Although measurement error models have been applied in air pollution^{35,36} and nutritional^{37,38} epidemiology, they have so far little been used in biomarker-based studies in the general population. Repeated measurements of exposure for each subject are needed to use these models; an option to limit the cost incurred by these repeated assessments is to perform them among a subsample of study subjects rather than the entire population.³⁸

Model Assumptions

Although no study actually tried to characterize the statistical nature (i.e., classical or Berkson-type errors) of the error entailed by within-subject biomarker variability, we believe that the assumption we made regarding the classical nature of error is plausible for biomarkers of exposure to widely used chemicals with short half-life in the general population. We assumed the correlation between two repeated measurements to be the same whatever the time elapsed between biospecimen collections, which might not be true for some short half-life compounds (correlation levels might for example decrease with time²¹). This might limit generalizability of our results. We assumed that all biomarker levels were above the limit of detection, which will generally not

occur for all populations nor chemicals considered with the currently available bioassays. The proportion of samples with a concentration below the limit of detection will be lower in pooled than unpooled samples,^{17,18} which may have caused us to overestimate the relative efficiency of measurement error models, compared with the pooling approach.

We assumed that the average of the concentrations measured in all biospecimens collected during the considered time window corresponded to the toxicologically relevant dose (or was a good proxy thereof). For some exposure–health outcome pairs, the toxicologically relevant measure of exposure may rather be the dose to a specific organ, of which the urinary concentration is only a proxy. These sources of exposure misclassification will possibly further bias dose–response relations in a way that we did not consider. Finally, we did not consider confounding and selection biases.

CONCLUSION

Biomarker-based studies dealing with compounds with an ICC of 0.6 or less can be strongly biased and weakly powered if only one biospecimen per subject is collected. Such studies should collect repeated biospecimens per subject. Assessing biomarker concentrations in each biospecimen allows, if measurement error models are used, to efficiently correct for both the between-assay error and the error related to the individual, but entails higher assay cost. The within-subject pooling approach that we described appears efficient in situations with low between-assay error and provides a less biased and more powerful design than if only one sample had been collected, without increasing assay costs. If the number of biospecimens pooled is identical for each participant, the pooling approach can be coupled with a posteriori bias disattenuation which, if a good estimate of the ICC is available for the studied chemical and under the assumption of a classical type error, further reduces bias.^{7,8}

ACKNOWLEDGMENTS

We acknowledge Irva Hertz-Picciotto for her support and Xavier Basagaña for a useful comment.

REFERENCES

- Calafat AM, Ye X, Silva MJ, Kuklennyk Z, Needham LL. Human exposure assessment to environmental chemicals using biomonitoring. *Int J Androl*. 2006;29:166–171; discussion 181.
- Schisterman EF, Albert PS. The biomarker revolution. *Stat Med*. 2012;31:2513–2515.
- Ye X, Wong LY, Bishop AM, Calafat AM. Variability of urinary concentrations of bisphenol A in spot samples, first morning voids, and 24-hour collections. *Environ Health Perspect*. 2011;119:983–988.
- Preau Jr JL, Wong LY, Silva MJ, Needham LL, Calafat AM. Variability over one week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among 8 adults: an observational study. *Environ Health Perspect*. 2010;118:1748–1754.
- Wielgomas B. Variability of urinary excretion of pyrethroid metabolites in seven persons over seven consecutive days: implications for observational studies. *Toxicol Lett*. 2013;221:15–22.
- Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med*. 1998;55:651–656.
- 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.
- Brunekreef B, Noy D, Clausing P. Variability of exposure measurements in environmental epidemiology. *Am J Epidemiol*. 1987;125:892–898.
- de Klerk NH, English DR, Armstrong BK. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. *Int J Epidemiol*. 1989;18:705–712.
- Bateson TF, Wright JM. Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *Am J Epidemiol*. 2010;172:344–352.
- Carroll RJ, Ruppert D, Stefanski LA. (Eds). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. London: Chapman & Hall; 1995.
- Braun JM, Kalkbrenner AE, Calafat AM, et al. Impact of early-life bisphenol A exposure on behavior and executive function in children. *Pediatrics*. 2011;128:873–882.
- Snijder CA, Heederik D, Pierik FH, et al. Fetal growth and prenatal exposure to bisphenol A: the generation R study. *Environ Health Perspect*. 2013;121:393–398.
- Tielemans E, Heederik D, Burdorf A, Loomis D, Habbema DF. Intraindividual variability and redundancy of semen parameters. *Epidemiology*. 1997;8:99–103.
- Dorfman R. The detection of defective members of large populations. *Ann Math Stat*. 1943;14:436–440.
- Malinovsky Y, Albert PS, Schisterman EF. Pooling designs for outcomes under a Gaussian random effects model. *Biometrics*. 2012;68:45–52.
- Weinberg CR, Umbach DM. Using pooled exposure assessment to improve efficiency in case-control studies. *Biometrics*. 1999;55:718–726.
- 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.
- Veronesi G, Ferrario MM, Chambless LE. Comparing measurement error correction methods for rate-of-change exposure variables in survival analysis. *Stat Methods Med Res*. 2013;22:583–597.
- Adibi JJ, Whyatt RM, Williams PL, et al. Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. *Environ Health Perspect*. 2008;116:467–473.
- 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.
- Philippat C, Botton J, Calafat AM, Ye X, Charles MA, Slama R; EDEN Study Group. Prenatal exposure to phenols and growth in boys. *Epidemiology*. 2014;25:625–635.
- Fuller WA. *Measurement Error Model*. New York, NY: John Wiley & Sons 1987.
- Hardin JW, Schmiediche H, Carroll RJ. The regression-calibration method for fitting generalized linear models with additive measurement error. *Stata J*. 2003;3:361–372.
- Hardin JW, Schmiediche H, Carroll RJ. The simulation extrapolation method for fitting generalized linear models with additive measurement error. *Stata J*. 2003;3:373–385.
- 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.
- Attfield KR, Hughes MD, Spengler JD, Lu C. Within- and between-child variation in repeated urinary pesticide metabolite measurements over a 1-year period. *Environ Health Perspect*. 2014;122:201–206.
- Slama R, Vrijheid M. Some challenges of studies aiming to relate the exposure to human health. *Occup Environ Med*. 2015;72:383–384.
- Patel CJ, Yang T, Hu Z, et al.; March of Dimes Prematurity Research Center at Stanford University School of Medicine. Investigation of maternal environmental exposures in association with self-reported preterm birth. *Reprod Toxicol*. 2014;45:1–7.
- Lenters V, Portengen L, Smit LA, et al. Phthalates, perfluoroalkyl acids, metals and organochlorines and reproductive function: a multipollutant

- assessment in Greenlandic, Polish and Ukrainian men. *Occup Environ Med.* 2015;72:385–393.
31. Braun JM, Kalkbrenner AE, Just AC, et al. Gestational exposure to endocrine-disrupting chemicals and reciprocal social, repetitive, and stereotypic behaviors in 4- and 5-year-old children: the HOME study. *Environ Health Perspect.* 2014;122:513–520.
 32. Christensen K, Sobus J, Phillips M, Blessinger T, Lorber M, Tan YM. Changes in epidemiologic associations with different exposure metrics: a case study of phthalate exposure associations with body mass index and waist circumference. *Environ Int.* 2014;73:66–76.
 33. Batistatou E, McNamee R. Performance of bias-correction methods for exposure measurement error using repeated measurements with and without missing data. *Stat Med.* 2012;31:3467–3480.
 34. Fung KY, Krewski D. Evaluation of regression calibration and SIMEX methods in logistic regression when one of the predictors is subject to additive measurement error. *J Epidemiol Biostat.* 1999;4:65–74.
 35. Van Roosbroeck S, Li R, Hoek G, Lebret E, Brunekreef B, Spiegelman D. Traffic-related outdoor air pollution and respiratory symptoms in children: the impact of adjustment for exposure measurement error. *Epidemiology.* 2008;19:409–416.
 36. Horick N, Weller E, Milton DK, Gold DR, Li R, Spiegelman D. Home endotoxin exposure and wheeze in infants: correction for bias due to exposure measurement error. *Environ Health Perspect.* 2006;114:135–140.
 37. Beydoun MA, Kaufman JS, Ibrahim J, Satia JA, Heiss G. Measurement error adjustment in essential fatty acid intake from a food frequency questionnaire: alternative approaches and methods. *BMC Med Res Methodol.* 2007;7:41.
 38. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr.* 1997;65(4 Suppl):1179S–1186S.