Relying on repeated biospecimens to reduce the effects of classical-type exposure measurement error in studies linking the exposome to health

Lydiane Agier, Rémy Slama, Xavier Basagaña

Relying on repeated biospecimens to reduce the effects of classical-type exposure measurement error in studies linking the exposome to health

Lydiane Agier[a*], PhD, Rémy Slama[a,#], PhD, Xavier Basagaña[b,c,d,#], PhD.

* corresponding author
[#] Equally contributed last authorship
[a] Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Inserm, CNRS, University Grenoble Alpes, Institute for Advanced Biosciences (IAB), U1209 Joint Research Center, Grenoble, France.
[b] ISGlobal, Barcelona, Spain.
[c] Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[d] CIBER Epidemiología y Salud Pública (CIBERESP), Spain.

Corresponding author:
Lydiane Agier
Institute for Advanced Biosciences
Inserm/CNRS/University Grenoble Alpes joint research center (U1209)
Rond-point de la Chantourne
38700 La Tronche
France
Tel : (0033) 4 76 54 94 00
Fax : (0033) 4 76 54 94 14
lydiane.agier@univ-grenoble-alpes.fr

Running head: Measurement error in exposome-health studies

Abbreviations : DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; LASSO: Least absolute shrinkage and selection operator; LOD: Limit of detection; MAB: Mean absolute bias for true predictors; Pooling: Within-subject biospecimens pooling; RC: Regression calibration; TP: True predictor.

ABSTRACT

The exposome calls for assessing numerous exposures, typically using biomarkers with varying amounts of measurement error, which can be assumed to be of classical type. We evaluated the impact of classical-type measurement error on the performance of exposome-health studies, and the efficiency of two measurement error correction methods relying on the collection of repeated biospecimens: within-subject biospecimens pooling and regression calibration. In a simulation study, we generated 237 exposures from a realistic correlation matrix, with various amounts of classical-type measurement error, and a continuous health outcome linearly influenced by exposures. Measurement error decreased the sensitivity to identify exposures influencing health from a value of 75% down to 46%, increased false discovery proportion from 26% to 49% and increased attenuation bias in the slope of true predictors from 45% to 66%. Assuming that repeated biospecimens were available, within-subject pooling and regression calibration improved sensitivity (which increased to 63%), false discovery proportion (down to 37%) and bias (down to 49%) compared to an error-prone study with a single biospecimen per subject. Performances were poorer for the exposures with the largest amount of measurement error, and increased with the number of available biospecimens. Relying on repeated biospecimens only for the exposures with the largest amount of measurement error provided similar performance improvement. Exposome studies relying on spot exposure biospecimens suffer from decreased performances if some biomarkers suffer from measurement error due to their temporal variability; performances can be improved by collecting repeated biospecimens per subject, in particular for non persistent chemicals.

Keywords: Biomarkers, exposome, measurement error, pooling, regression calibration.

Funding: This work was supported by the European Union: the Seventh Framework Programme (FP7/2007-206) for the HELIX project [grant number 308333 to Martine Vrijheid]; and Horizon 2020 Research and Innovation Programme for the HBM4EU project [grant number 733032 to Marike Kolossa-Gehring]. HBM4EU represents a joint effort of 28 countries, the European Environment Agency and the European Commission, co-funded by Horizon 2020. The main aim of the initiative is to coordinate and advance human biomonitoring in Europe. HBM4EU provides evidence of the actual exposure of citizens to chemicals and the possible health effects to support policy making. The project involves collaboration between several Commission services, EU agencies, national representatives, stakeholders and scientists, demonstrating how research funding can build bridges between the research and policy worlds. HBM4EU launched on 1 January 2017 and will run until 31 December 2021. HBM4EU will contribute to the overarching goal of ensuring that EU chemical policies minimize the adverse effects of chemicals on human health. We acknowledge the support of Région Auvergne Rhône-Alpes for scientific collaborations with Catalogna.

Most of the computations presented in this paper were performed using the CIMENT infrastructure (https://ciment.ujf-grenoble.fr), which is supported by the Auvergne Rhône-Alpes region (grant CPER07_13 CIRA: http://www.ci-ra.org). The support of SCUSI program from Région Auvergne Rhône-Alpes is also acknowledged.

INTRODUCTION

The exposome concept was defined as encompassing the totality of environmental exposures from the prenatal period onwards (Wild, 2012). Exposome studies typically rely on biomarkers to assess chemical exposure from multiple sources (Calafat et al., 2006; Schisterman and Albert, 2012); often, a single biospecimen is collected in each subject (LaKind et al., 2019; Perrier et al., 2016). For chemicals with a short biological half-life, such as bisphenol A or phthalates, within-subject biomarker concentrations strongly vary over time. Whatever the accuracy of the measurement technique used, this will induce measurement error, mainly for the least persistent compounds (those with the lowest intraclass coefficient of correlation, or ICC), for which a spot biospecimen will provide a poor estimate of exposure over long time windows (Casas et al., 2018; Perrier et al., 2016; Preau et al., 2010; Vernet et al., 2019; Wielgomas, 2013; Ye et al., 2011). A realistic assumption is that chemicals exposure is measured by biomarkers with an independent additive error, such that the within-subject average of repeated measurements is an unbiased estimate of exposure. This type of measurement error corresponds to classical- type measurement error (Carroll et al., 2006). In a single exposure context, classical-type exposure measurement error can strongly impact the estimation of exposure-health relationships (Armstrong, 1998; Brakenhoff et al., 2018a; Carroll et al., 2006; de Klerk et al., 1989; Jurek et al., 2005; Perrier et al., 2016): naïve models not accounting for measurement error provide regression estimates that are attenuated, and have decreased statistical power compared to the ideal situation without measurement error (Carroll et al., 2006). Several statistical techniques have been developed to limit estimation bias when classical type exposure measurement error is present. Regression calibration (RC) (Carroll et al., 2006) is often used, and provides approximately consistent estimates (Buzas et al., 2005). It requires either a (within-subject) repeated assessment of

exposures, at least in a subset of the population, or an unbiased estimate of exposures' ICC. An alternative, also relying on the collection of repeated biospecimens, is to pool samples within-subject and assess the exposures biomarker in the pooled sample, corresponding to the so-called *within-subject biospecimens pooling approach* (Perrier et al., 2016; Vernet et al., 2019). In a single exposure context, within-subject biospecimens pooling, possibly followed by a posteriori disattenuation of coefficients estimates, is efficient to limit attenuation bias in the exposure-health association and to increase power (Perrier et al., 2016; Vernet et al., 2019).

The impact of classical-type exposure measurement error on bias and power_may be compounded in an exposome context. Indeed, in exposome epidemiological studies, efficiently identifying the exposures affecting health (i.e. with good sensitivity, or power; and low false detection proportion, or FDP) is generally a challenge, in particular when exposures show some correlation (Agier et al., 2016), which is a realistic assumption (Tamayo-Uria et al., 2019). Additionally, exposome studies can simultaneously consider several biomarker-based exposures with differential measurement error (i.e. of different amplitudes across exposures) (Slama and Vrijheid, 2015); typically, persistent pollutants such as PCBs, for which a spot biospecimen provides a rather good exposure proxy, and non persistent chemicals, such as organophosphate pesticides or bisphenol A, with ICCs based on spot biospecimens in the 0.1-0.5 range (Casas et al., 2018), may both be included in the same study (Agier et al., 2020, 2019). This differential measurement error in exposome studies is likely to further affect the efficiency to identify exposures affecting health, beyond the performance loss due to the consideration of a large number of correlated compounds. Indeed, when applying a variable selection model in an exposome study relying on spot biospecimens, the chances of a given exposure being selected might be lower for compounds with a low ICC (Perrier et al., 2016; Vernet et al., 2019). In addition, in a multivariable situation, regression

estimates can be biased in any direction as a consequence of classical-type measurement error (Brakenhoff et al., 2018b; Carroll et al., 2006).

Up to date, three simulation studies have estimated the performances of specific variable selection methods in assessing exposure-outcome associations in a multi-exposure context with classical-type exposure measurement error. Guangning et al. tested several two-step approaches (applying a measurement error correction method and using the resulting exposure proxies in a variable selection method) in a 2-exposures context (Guangning, 2014). Vasquez et al. applied a corrected least absolute shrinkage and selection operator (LASSO) procedure to 100 uncorrelated exposures (Vasquez et al., 2019). Brown et al. compared LASSO to a novel variable selection method in a context where 10 exposures affected the outcome, were not correlated to the other 90 exposures, and jointly explained over 90% of the outcome variability (Brown et al., 2019), which is much higher than the outcome variability that a few environmental exposures are expected to explain. To our knowledge, no study assessed the impact of exposure measurement error on exposure-outcome relationship in a realistic exposome context, i.e. in a context with many exposures that are all correlated (Tamayo-Uria et al., 2019), with a small global explanatory power of exposures on the outcome. Moreover, the efficiency of the recently proposed within-subject biospecimens pooling approach (Perrier et al., 2016) has never been estimated in a multi-exposure context.

Using a simulation approach relying on a realistic exposome correlation structure, we aimed to assess the performances of RC and of within-subject biospecimens pooling in handling differential classical-type exposure measurement error in exposome-health studies.

MATERIALS AND METHODS

1. Overview of the simulation model

We simulated an exposome study with p=1, 10 or 237 error-prone exposures in a fictitious population of 1 200 subjects (the approximate population size of the HELIX exposome project, which we mainly refer to for the design of this simulation study (Agier et al., 2020; Vrijheid et al., 2014)). We assumed that k=0, 1, 3 or 10 exposures linearly influenced a health outcome **Y**, which is an assumption of rare (and weak, see below) features realistic for 'omics data, including possibly the exposome (Donoho and Jin, 2008). We considered that all exposures suffered classical-type measurement error, with varying levels of error which were quantified from their intra-class correlation coefficient (ICC). Exposure-health associations were assessed using 2-step approaches. We first applied either RC or within-subject biospecimens pooling, assuming that 2, 5 or 10 error-contaminated biospecimens (collecting 10 biospecimens per subject is now logistically feasible (Lyon-Caen et al., 2019)) were available per subject (we did not cover a greater range of values for the number of biospecimens per subject because our aim is to illustrate that repeated biospecimens collection is a solution, without necessarily trying to identify an optimum number of biospecimens per subject). With both these methods, we obtained a set of transformed exposures over which, in the second step, we applied a variable selection method (Deletion/Substitution/Addition algorithm, DSA, (Sinisi and van der Laan, 2004)) to identify exposures associated with **Y**. The performances of the RC and pooling approaches were quantified through their sensitivity, bias for causal exposures, and false discovery proportion, which were averaged over 200 datasets (or simulation runs). Performances were compared to those obtained when directly applying DSA either on the error-free exposures (which corresponds to the ideal approach); or on a dataset including a single error-contaminated measure of each exposure per individual, first without applying any

measurement error correction (named hereafter the *naive approach,* what epidemiologists ignoring measurement error would do, and which is expected to correspond to the least efficient approach), and second after applying RC using *a priori* known ICC values for all exposures.

2. Generation of exposures

We generated a matrix $\boldsymbol{T} = (T_i)_{\mathbf{i=1,\ldots,p}}$ of p error-free exposures from a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$ (i.e. $\boldsymbol{T} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$), in a fictitious population of N=1 200 subjects. $\boldsymbol{\Sigma}$ was centered-reduced such that all exposures were generated with an average of 0 and a variance of 1, and could easily be compared. For p=237 exposures, $\boldsymbol{\Sigma}$ was obtained from environmental factors measured in the INfancia y Medio Ambiente mother-child, or INMA cohort (one of the HELIX cohorts, Appendix A) (Guxens et al., 2012). For p=10 exposures, $\boldsymbol{\Sigma}$ was generated so as to include a mixture of low and high pairwise correlations (in order to cover the range of correlation levels observed for p=237). For p=1 exposure, its variance was also set to 1 ($\boldsymbol{\Sigma}$=1). .

Classical-type error assumes that the exposures are measured with independent additive errors. We thus generated, for each exposure T_i, 10 error-contaminated biospecimens per individual, denoted $(W_{ij})_{\mathrm{j=1,\ldots,10}}$, that were assumed to be collected at random time points within the exposure window of interest, using the same equation as in (Perrier et al., 2016):

$$W_{ij} = T_i + U_{ij} \qquad \text{Eq. (1)}$$

where $U_{ij} \sim \mathcal{N}(0,\ \sigma^2_{U_i})$, with $\sigma^2_{U_i} = \sigma^2_{T_i}\left(\frac{1}{ICC_i} - 1\right)$, $\sigma^2_{T_i}$ being the variance and ICC_i the ICC of exposure T_i. We generated ICC_i in the [0.15, 1.00] range, from a balanced mixture of two normal distributions $\mathcal{N}(0.95, 0.15)$ and $\mathcal{N}(0.50, 0.20)$ (Appendix B).

3. Generation of the health outcome

The health outcome Y was generated as a function of the error-free exposures according to:

$$Y = \sum_{i=1}^{p} \beta_i T_i + \epsilon \qquad \text{Eq.(2)}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and where the regression coefficients β_i were set to 0, except for the k true predictors (TP) that were assumed to be causally related to the outcome, for which β_i was 1. TPs were randomly selected amongst all exposures at each simulation run. The proportion of variance explained by the k TPs was set to $R^2 = 3\% \times k$.

4. Estimation of the exposure-health association

Two methods aiming to correct for exposure measurement error were applied to the simulated data. In the within-subject biospecimens pooling approach (Perrier et al., 2016), all concentrations W_{ij} that are considered available for each exposure T_i are averaged at the individual level with equal weights; this measure is named hereafter $\widehat{\boldsymbol{T}}_{\boldsymbol{pool}} = \overline{\boldsymbol{W}} = \left(\overline{\boldsymbol{W_i}}\right)_{\boldsymbol{i=1,\ldots,p}}$. In the RC approach (Carroll et al., 2006), in the absence of adjustment factors, proxies of exposures are obtained by predicting $\boldsymbol{T} = (T_i)_{i=1,\ldots,p}$ from all available error-contaminated measures relying on a multivariate linear regression model:

$$\widehat{\boldsymbol{T}}_{\boldsymbol{RC}} = \hat{\mu}_{\overline{W}} + \frac{\widehat{\boldsymbol{\Sigma}}_{\mathbf{T}}^{\,t}}{\widehat{\boldsymbol{\Sigma}}_{\mathbf{T}} + \widehat{\boldsymbol{\Sigma}}_{\mathbf{U}}} (\overline{W} - \hat{\mu}_{\overline{W}}) \qquad \text{Eq. (3)}$$

where $\hat{\mu}_{\overline{W}}$ is the exposure-by-exposure mean of $\overline{\boldsymbol{W}}$, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{T}}$ is the estimated covariance matrix of the error-free exposures $\boldsymbol{T}$, and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{U}}$ is the estimated covariance matrix of measurement error terms, with diagonal coefficients being divided by the average number of repeated biospecimens per subject for the given exposure. The estimated ICCs (named $(\widehat{ICC}_i)_{i=1,\ldots,p}$)

can be derived from $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}$. When no repeated biospecimens are available, one has to assume that the ICCs for the given exposures are known from a previous study.

We separately assessed the associations of $\widehat{\boldsymbol{T}}_{\boldsymbol{pool}}$ and $\widehat{\boldsymbol{T}}_{\boldsymbol{RC}}$ with the health outcome Y using the deletion/substitution/addition (DSA) algorithm (Sinisi and van der Laan, 2004). We selected this sequential variable selection algorithm for it was shown to have equivalent or better performances in comparison with other dimension reduction or variable selection approaches such as elastic net or exposome-wide association study in a simulation study based on a similar approach and similar hypotheses but without exposure measurement error (Agier et al., 2016). DSA model search process starts with the intercept model, and, at each iteration, considers removing a term, replacing a term by another, or adding a term to the current model. The final model is selected by minimizing the prediction root mean squared error using 5-fold cross-validated data (see Appendix C for considerations on the selection stability of DSA). We did not allow polynomial or interaction terms, and limited the maximum model size to 40 covariates, a number never reached in our simulations.

5. Statistical performance assessment

For each simulation run and scenario (defined by the number of exposures, the exposures correlation structure, the number of TP, and the number of biospecimen available per exposure and per subject), we assessed the performances of each approach by calculating the sensitivity, defined as the proportion of TP that were selected by the method; and the false discovery proportion (FDP), defined as the proportion of selected variables that were not TP. When no variable was selected, the FDP was given a value of 0. We also computed the mean absolute bias for TP, which measures the accuracy of the estimated coefficients by comparing the coefficient value β_i that was used to generate the outcome Y with its estimated value $\hat{\beta}_i$ obtained by DSA, i.e. :

$$\frac{1}{k}\sum_{i=1}^{p}|\beta_i - \hat{\beta}_i| \times \beta_i \qquad \text{Eq. (4)}$$

(note that β_i only takes values 0 or 1).

Performances were averaged by scenario and measurement error correction method. We further investigated the effect of the ICC level by stratifying exposures by ICC decile and estimating these criteria values within each decile group.

6. Additional scenarios

Unbalanced designs: In order to limit the number of biospecimens to be collected, we considered: (i) collecting repeated biospecimens in a subset of the study participants only, i.e. in 10%, 40%, or 70% randomly selected subjects within our population; or (ii) collecting or assaying repeated biospecimens only for exposures known to have a high within-subject variability, i.e. exposures with an ICC lower than 0.4, 0.6 or 0.8 (in this case, RC requires an *a priori* known ICC value for each exposure without repeated biospecimens).

Varying correlation levels: Three additional situations were considered with varying exposures correlation structures: one in which the off-diagonal elements of the covariance matrix of the true exposures $\mathbf{\Sigma}$ were divided by two (labelled $\mathbf{\Sigma}^{-}$); one in which they were multiplied by two (labelled $\mathbf{\Sigma}^{+}$) and one with a diagonal correlation structure ($\mathbf{\Sigma}^{0}$). When needed, we computed the closest positive semi-definite matrix of the correlation structure before generating the exposures (Agier et al., 2016).

These scenarios were tested in a p=237 exposures context only, with 100 simulation runs.

Analyses were performed using the R software (www.r-project.org). Computing code is provided in Appendix D.

RESULTS

1. Low dimension exposure-health studies

In an exposure-health study assuming that 10 exposures were assessed without measurement error and that a single TP affected the health outcome, sensitivity was 100%, mean absolute bias for TP was 0.15, and FDP was 18%. In the presence of uncorrected measurement error, sensitivity decreased to 94%, bias increased to 0.42, and FDP increased to 31%. Applying measurement error correction methods improved the performances compared to this naive approach, with efficiency raising as the number of repeated biospecimens increased and reaching, with both methods, values that were similar to the case without exposure measurement error (Table 1). Similar results were obtained when we considered 3 or 10 TP, or in a single exposure-health study (Appendix E).

Exposure assessment	Measurement error correction	1-exposure study [1]		10-exposures study			237-exposures study		
		Sensitivity	MAB [2]	Sensitivity	FDP	MAB [2]	Sensitivity	FDP	MAB [2]
Error-free exposures	None	1.00	0.13	1.00	0.18	0.15	0.80	0.24	0.30
Error-contaminated biospecimens									
1 biospecimen	None (naive approach)	0.98	0.41	0.94	0.31	0.42	0.49	0.40	0.63
	RC (ICCs provided)	0.98	0.18	0.90	0.24	0.26	0.56	0.41	0.52
2 biospecimens	Pooling	0.99	0.29	0.98	0.24	0.31	0.60	0.39	0.53
	RC	0.99	0.15	0.96	0.22	0.19	0.62	0.37	0.47
5 biospecimens	Pooling	1.00	0.18	1.00	0.19	0.21	0.68	0.34	0.43
	RC	1.00	0.14	0.98	0.17	0.17	0.76	0.26	0.34
10 biospecimens	Pooling	1.00	0.15	1.00	0.22	0.18	0.72	0.31	0.38
	RC	1.00	0.13	0.98	0.18	0.17	0.77	0.25	0.33

Table 1. Performances in identifying exposure-health associations, scenarios considering either 1, 10, or 237 exposures in total.

Sensitivity, FDP and mean absolute bias for true predictors (MAB) are estimated, considering scenarios with the continuous health outcome being affected by a single true predictor that was randomly selected amongst all available exposures at each simulation run. Results are given for DSA directly applied on the error-free exposures and on one error-contaminated measure (naive approach); for DSA after applying RC on one error-contaminated measure with ICCs values being provided; and for DSA after applying RC or the within-subject biospecimens pooling approach on error-contaminated measures with a balanced design with 2, 5 or 10 repeated measures per subject. Average values computed over 200 simulation runs are displayed. See Figure 1 and Appendix F for results with 237 exposures and more than 1 true predictor, Appendix E for results with 10 exposures and more than 1 true predictor.

Abbreviations: DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; MAB: Mean absolute bias for true predictors (variables that were not predictors were ignored); Pooling: Within-subject biospecimens pooling; RC: Regression calibration.

[1] FDP is not defined when there is a single exposure in the study and it is a true predictor (which means there is no exposure that is not a true predictor).

[2] MAB is estimated among true predictors only, for which the true parameter value β_i was one.

2. Exposome-health studies

When we considered an exposome context with 237 exposures and a single TP affecting the health outcome, in the absence of measurement error, sensitivity was 80%, FDP was 24% and bias for TP was 0.30. In the presence of measurement error, if no repeated biospecimen was available, the naive approach resulted in a strong performance deterioration: sensitivity decreased to 49%, FDP was 40% and bias for TP was 0.63. Assuming that external ICCs were available, RC allowed slightly improving sensitivity (56%) and bias (0.52), while FDP remained unchanged. When repeated biospecimens were available and a balanced design was considered (i.e. an equal number of repeated biospecimens for all subjects and exposures), performances of methods correcting for measurement error varied between those of the naive approach and those obtained in the absence of measurement error (without reaching the latter). RC was marginally more efficient than pooling, with a major improvement being observed when increasing the number of repeated biospecimens from 2 to 5, and almost no difference between the scenarios with 5 and 10 repeated biospecimens. In comparison, pooling displayed a more regular improvement in performances as the number of repeated biospecimens increased (Figure 1, Table 1). Similar observations were made when we considered 3 or 10 TP, except that performances were slightly decreased for all scenarios (an overall 13% decrease in sensitivity, 11% increase in FDP and 0.19 increase in bias when considering 10 TP compared to 1 TP, Figure 1). When we assumed that no exposure affected the health outcome, the chance to mistakenly select an exposure that was not a TP did not increase as a result of exposures measurement error, compared to a situation where exposures were measured without error (Appendix G).

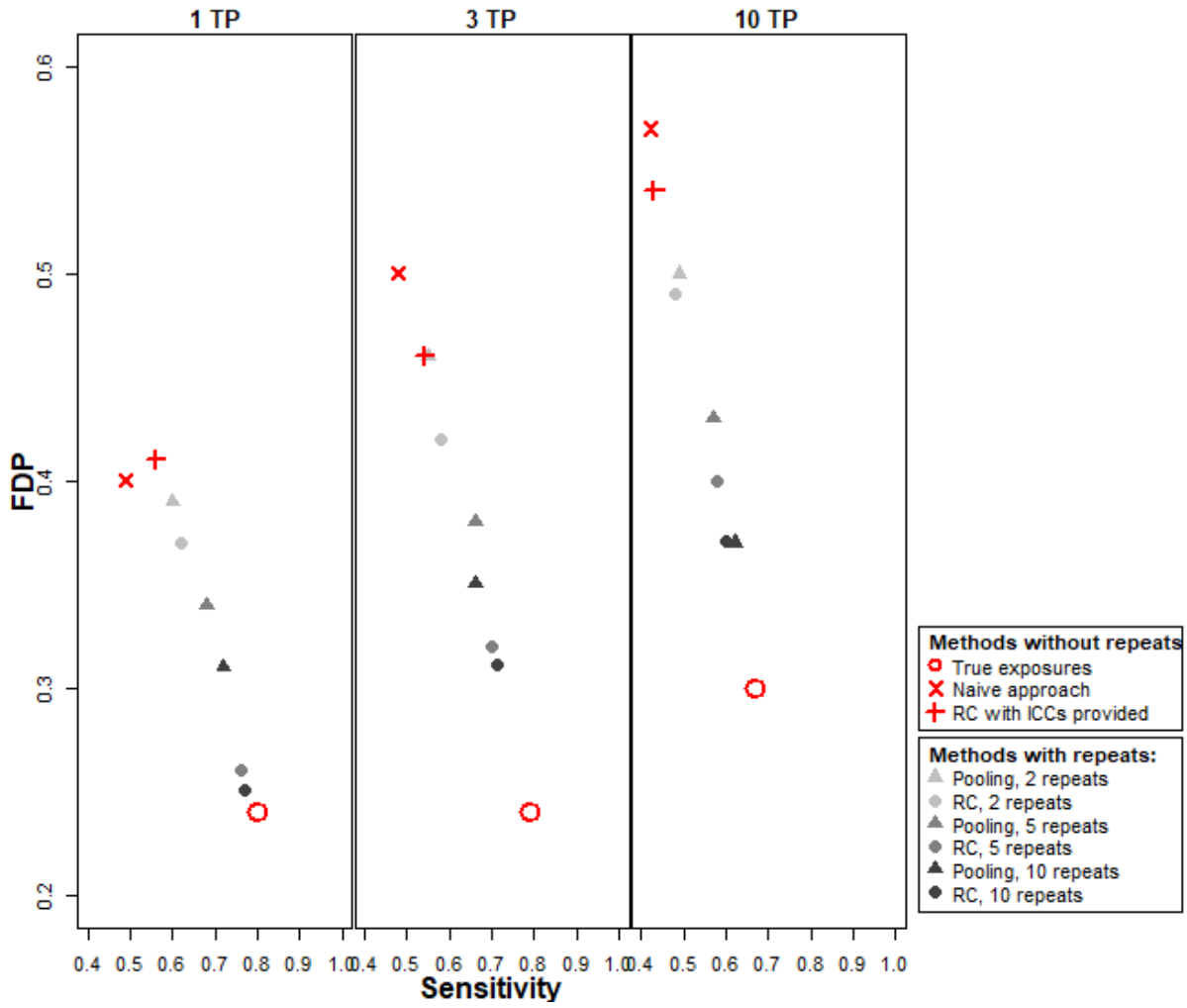

Figure 1. Sensitivity and FDP, considering scenarios with the continuous health outcome being affected by 1, 3 or 10 true predictors randomly selected amongst a set of 237 exposures.

Results are given for DSA directly applied on the error-free exposures (red circle) and on one error-contaminated measure (naive approach, red cross); for DSA after applying RC on one error-contaminated measure with ICCs values being provided (red + sign); and for DSA after applying RC (triangles) or the within-subject biospecimens pooling approach (circles) on error-contaminated measures with a balanced design with 2, 5 or 10 repeated measures per subject. Average values computed over 200 simulation runs are displayed (see also Appendix F).

DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; Pooling: Within-subject biospecimens pooling; RC: Regression calibration; TP: True predictors.

Plotting performances as a function of the exposures' ICC (Figure 2) showed that for persistent exposures (those with an ICC close to 1), sensitivity and bias were close to the ones

observed in the absence of measurement error, even when using the naïve approach. The performances of both methods decreased as the exposure variability increased (i.e. when the ICC decreased), with a linear decline in sensitivity and a linear increase in bias (Figure 2 A, C). In contrast, the risk of false discovery (FDP, Figure 2B) was always greater than the one observed in the absence of measurement error, and varied little as a function of the ICC in the [0.4, 1.0] range: for a given number of repeated biospecimens and a given correction method (including no correction), all exposures with an ICC in the [0.4, 1.0] range had a similar chance to be mistakenly selected when they were not truly associated with the outcome. In comparison, the least persistent exposures (those with an ICC in the [0.15;0.40] range) had a high risk of being mistakenly selected, even when 5 or 10 repeated biospecimens were collected.

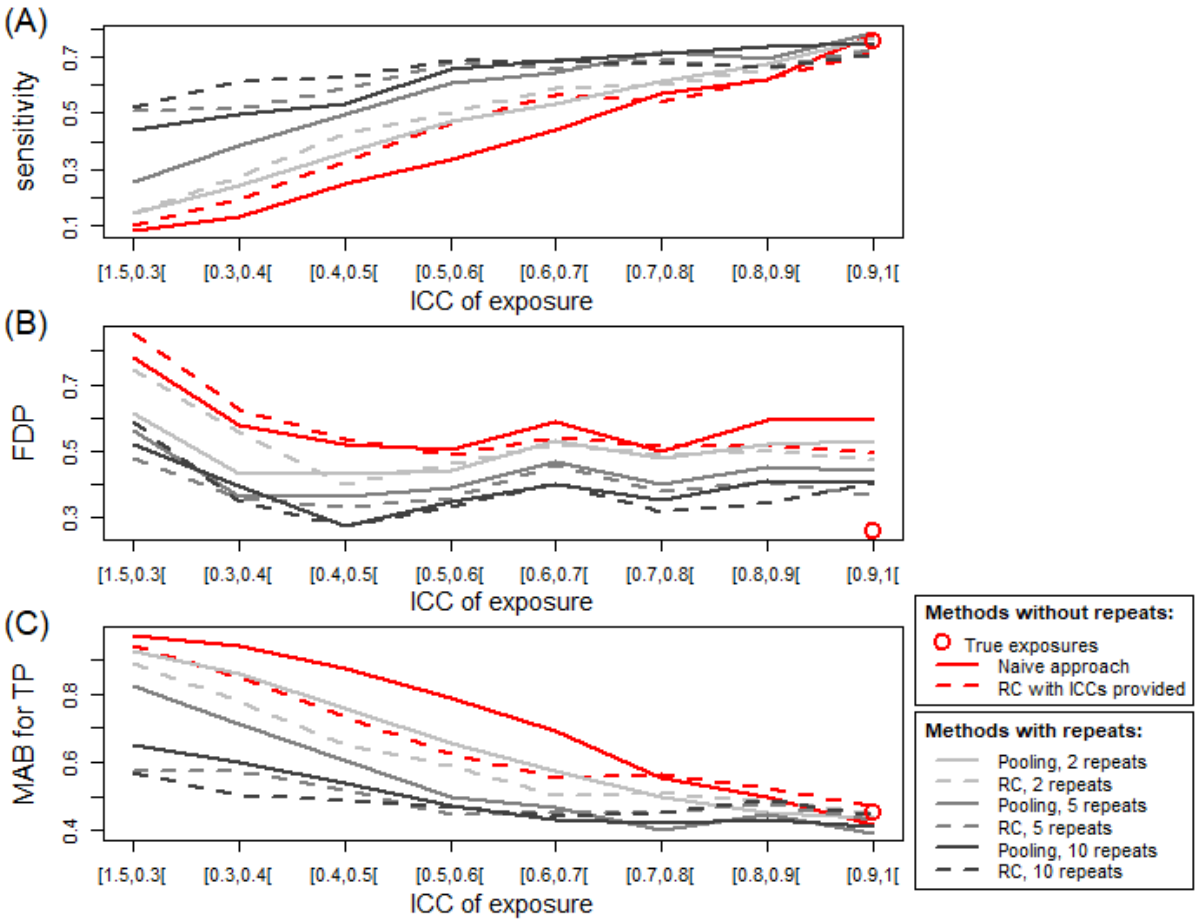

Figure 2. Sensitivity (A), FDP (B) and mean absolute bias for true predictors (C) amongst exposures inside each ICC decile, considering scenarios with the continuous health outcome being affected by 1, 3 or 10 true predictors randomly selected amongst a set of 237 exposures.

Results are given for DSA directly applied on error-free exposures (red circle) and on one error-contaminated measure (naive approach, plain thick line); for DSA after applying RC on one error-contaminated measure with ICCs values being provided (dashed thick line); and for DSA after applying RC (plain lines) or the within-subject biospecimens pooling approach (dashed lines) on error-contaminated measures with a balanced design with 2, 5 or 10 repeated measures per subject. Average values per ICC decile computed over all simulation runs (200 simulation runs per scenario, pooled results) are displayed.

DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; MAB: Mean absolute bias (variables that were not predictors were ignored); Pooling: Within-subject biospecimens pooling; RC: Regression calibration; TP: True predictors.

3. Unbalanced designs

The performances of the methods decreased linearly as the proportion of subjects with repeated biospecimens decreased from 100% to 40%; at 40% of subjects with repeated biospecimens, performances were similar to the ones observed with the naive approach; at 10% of subjects with repeated biospecimens, measurement error correction methods displayed even some loss over the naive approach (Figure 3). RC performed better than the pooling approach, except when 2 repeated biospecimens were collected for less than half of the subjects. Sampling 5 biospecimens in 10% of the population, or 2 biospecimens in 40% of the population, the remaining subjects having a unique exposure measure (which corresponds to the same total number of biospecimens collected and assessed, 1.4 times the number of subjects) made no difference in terms of performance in capturing the exposure-health association (Appendix H). Repeated biospecimens collection could be restricted to exposures with an ICC lower than 0.6 (corresponding to 42% of exposures) without losing the benefits of measurement error correction methods (Figure 4); that is, performances were similar to those obtained when collecting repeated biospecimens for all exposures. Restricting repeated

biospecimens to exposures with an ICC below 0.4 inflicted a substantial performance decrease.

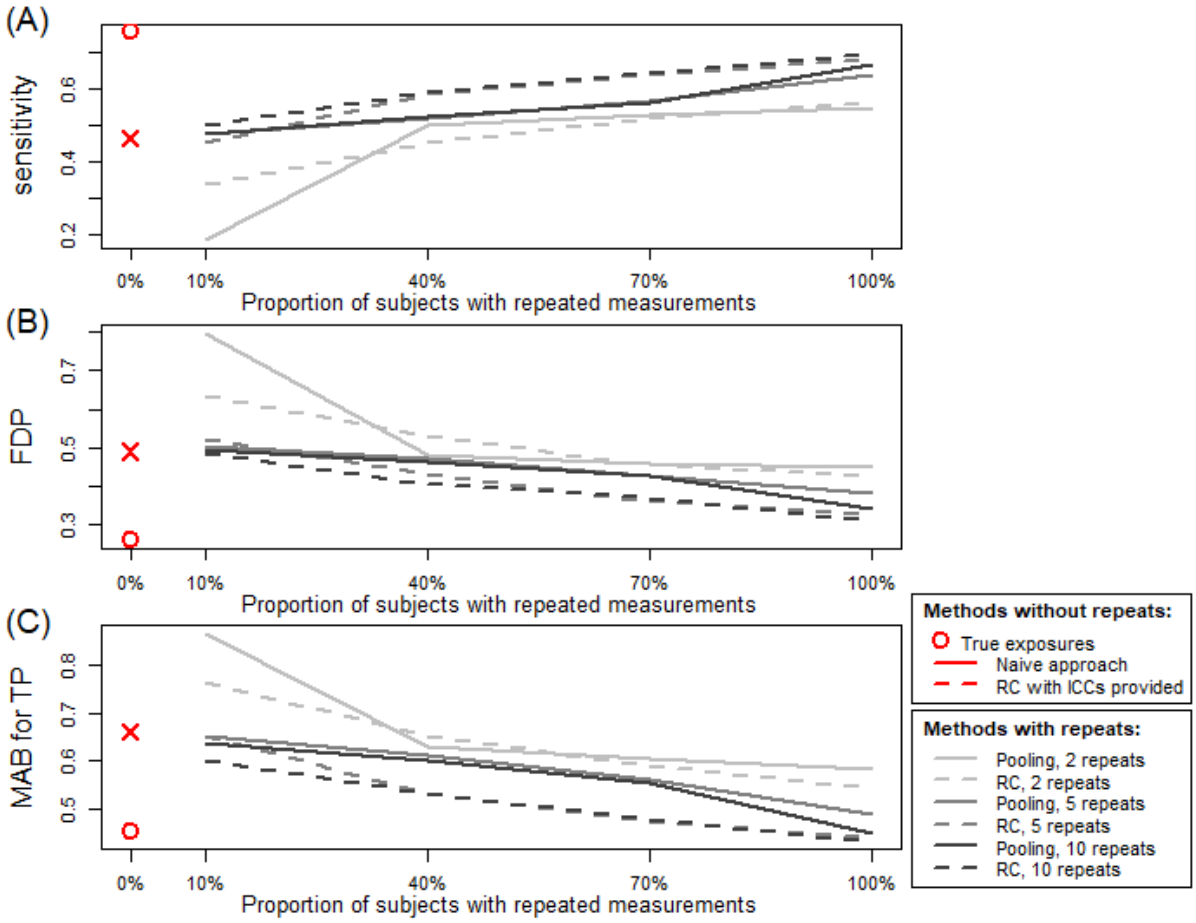

Figure 3. Sensitivity (A), FDP (B) and mean absolute bias for true predictors (C) when repeated biospecimens were collected in a subset of subjects, considering scenarios with the continuous health outcome being affected by 1, 3 or 10 true predictors randomly selected amongst a set of 237 exposures.

We considered the proportion of subjects with repeated biospecimens increased from 10% to 100%. Results are given for DSA directly applied on error-free exposures (red circle) and on one error-contaminated measure (naive approach, x symbol); and for DSA after applying RC (plain lines) or the within-subject biospecimens pooling approach (dashed lines) on error-contaminated measures with an unbalanced design with 2, 5 or 10 repeated measures per subject being collected in a subgroup of the study population, a single biospecimen being collected otherwise. Average values computed over all simulation runs (100 simulation runs per scenario, pooled results) are displayed.

DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; MAB: Mean absolute bias (variables that were not predictors were ignored); Pooling: Within-subject biospecimens pooling; RC: Regression calibration; TP: True predictors.

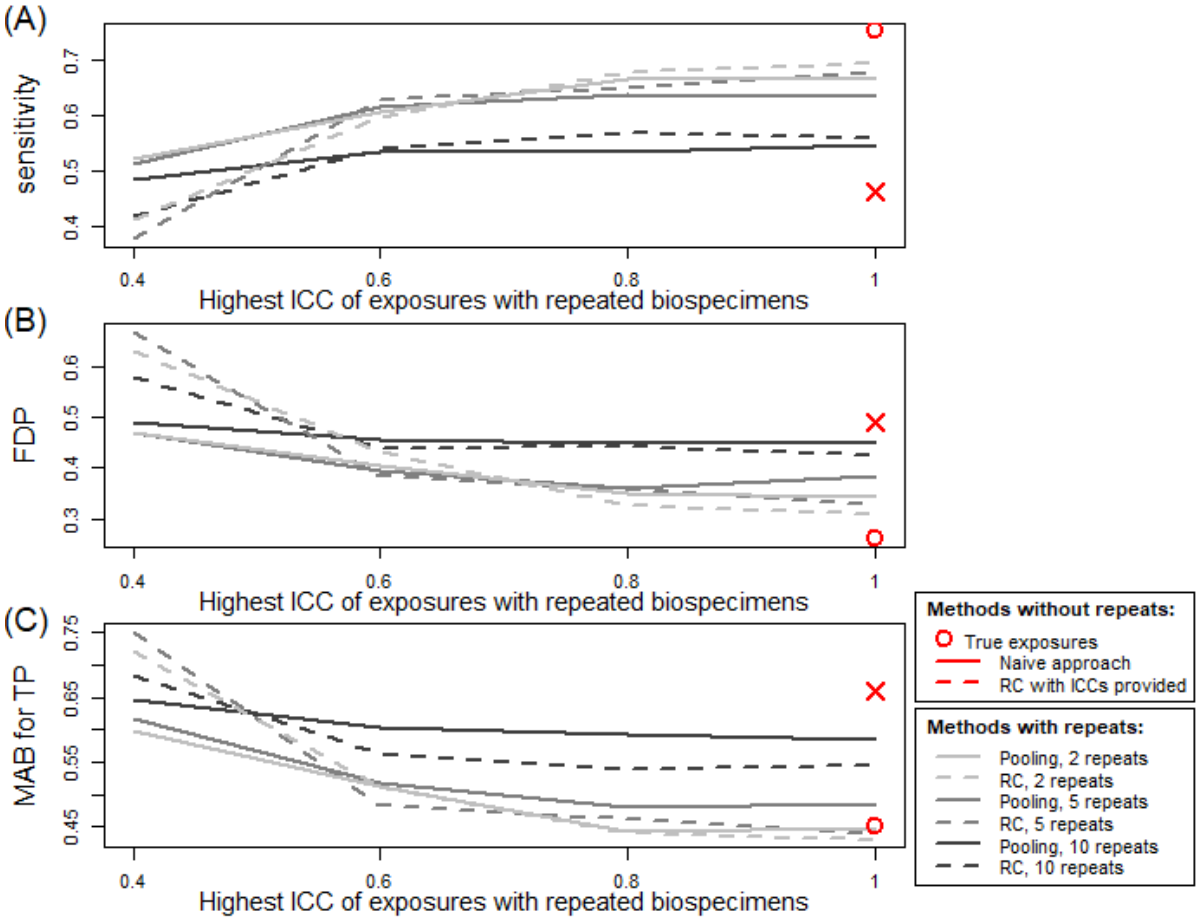

Figure 4. Sensitivity (A), FDP (B) and mean absolute bias for true predictors (C) when repeated biospecimens were collected in a subset of exposures, considering scenarios with the continuous health outcome being affected by 1, 3 or 10 true predictors randomly selected amongst a set of 237 exposures.

We considered values for the ICC threshold above which no repeated biospecimen is collected increased from 0.4 to 1 (1 meaning that repeated biospecimens were collected for all exposures). Results are given for DSA directly applied on error-free exposures (red circle) and on one error-contaminated measure (naive approach, x symbol); and for DSA after applying RC (plain lines) or the within-subject biospecimens pooling approach (dashed lines) on error-contaminated measures with an unbalanced design with 2, 5 or 10 repeated measures per subject being collected in a subset of exposures, a single biospecimen being collected otherwise. Average values computed over all simulation runs (100 simulation runs per scenario, pooled results) are displayed.

DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion; ICC: Intra-class correlation; MAB: Mean absolute bias (variables that were not predictors were ignored); Pooling: Within-subject biospecimens pooling; RC: Regression calibration; TP: True predictors.

4. Exposures correlation levels

The degree of correlation amongst exposures had a large impact on the models' performances (Appendix I). When relying on an existing exposome correlation structure (median [1st quartile, 3rd quartile] absolute value of the coefficients of correlation, 0.06 [0.03, 0.15]), over all scenarios investigated (including scenarios with error-free exposures), average (range) sensitivity was 61% (42%;80%) and average FDP was 38% (24%;57%). When correlation levels amongst exposures were double the initial values ($\boldsymbol{\Sigma}^{+}$ correlation matrix), performances diminished: sensitivity decreased to 37% (16%; 66%) and FDP increased to 62% (41%; 78%). With correlation levels amongst exposures half the initial values ($\boldsymbol{\Sigma}^{-}$ correlation matric), performances were higher: sensitivity was 86% (64%; 98%) and FDP was 14% (6%; 31%). Variable selection methods were almost perfectly efficient when exposures were uncorrelated: sensitivity was 90% (66%; 100%) and FDP was 6% (2%; 11%). The impact of the number of repeated biospecimens and of TPs on the performances was similar across all correlation matrices tested.

DISCUSSION

To our knowledge, this study is the first to describe the impact of exposure measurement error in realistic exposome-health studies. The decrease of the performance of variable selection models due to measurement error is known to be substantial in low-dimension settings (Perrier et al., 2016; Vernet et al., 2019). We showed that this phenomenon is compounded in an exposome context, with increased false positive rates, false negative rates, and bias in dose-response functions; and that these problems aggravate when correlation within the exposome increases. Compared to single exposure studies of similar population size without measurement error, exposome studies ignoring issues related to measurement error suffer from a double source of performance loss, due to the increase in the number of exposures considered (and to the resulting exposures correlation), and to measurement error (Figure 5). The performance loss due to measurement error can in large part be recovered by collecting repeated biospecimens (the more biospecimens, the better the performances) and using methods such as regression calibration or within-subject biospecimens pooling. As expected, the impact of measurement error was larger for compounds with the largest variability, so that, for a given strength of association and assuming lack of other biases, exposome studies are more likely to correctly identify the effect of the most persistent compounds, compared to the least persistent ones – an issue that could be termed ICC-related differential sensitivity. On the other hand, most exposures, except the least persistent ones, had a similar risk of being mistakenly selected when they were not associated with the outcome. Limiting repeated samples to the least persistent exposures is an option to limit costs without affecting variable selection performances.

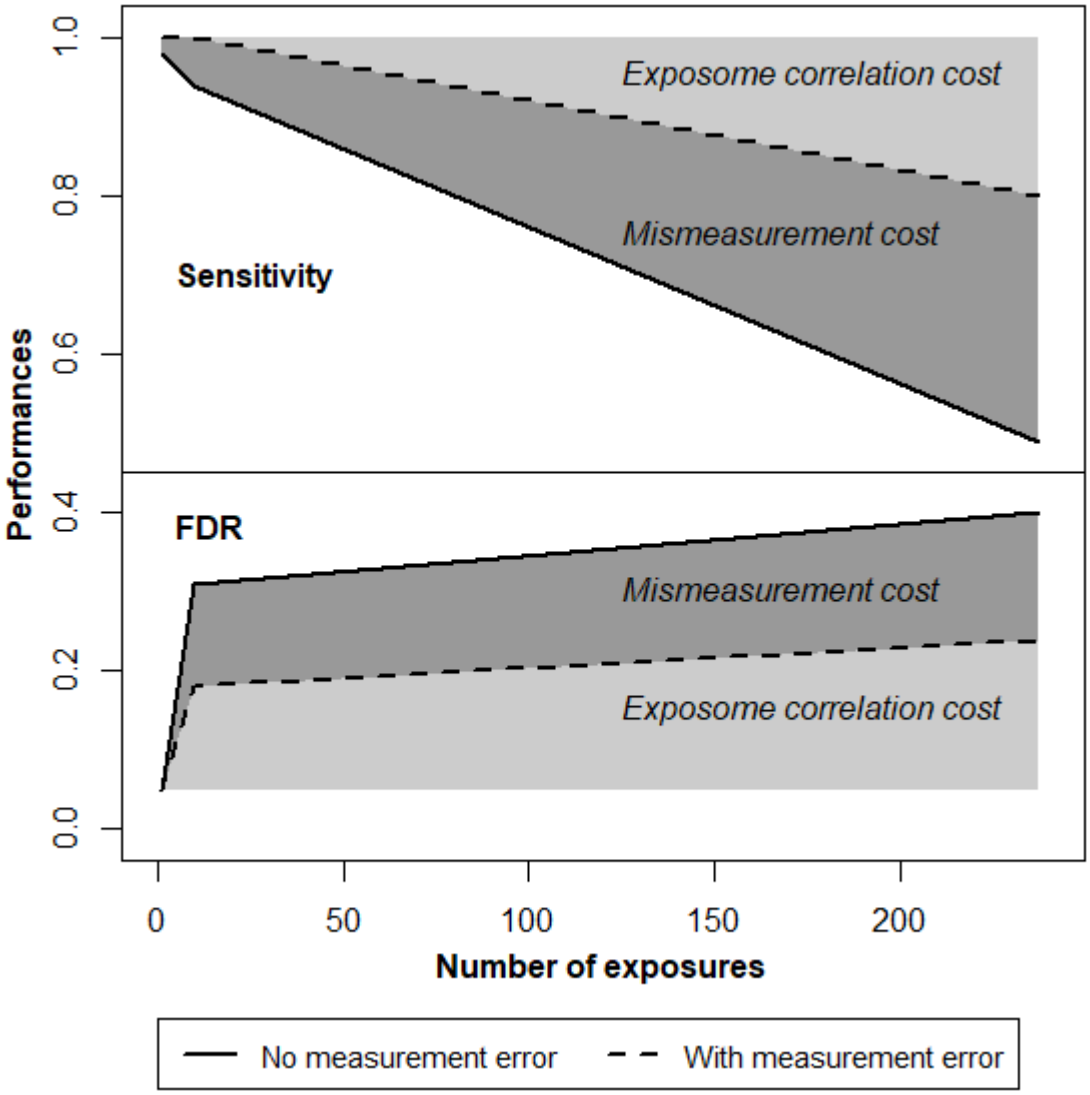

Figure 5. Sensitivity and FDP loss in exposome studies due to an increase in the number of exposures considered (the exposome correlation cost) and to exposures measurement error (the mismeasurement cost).

We considered scenarios with the continuous health outcome being affected by a single true predictor randomly selected amongst a set of 1, 10 or 237 exposures. Results are given for DSA directly applied on the available data (in the presence of measurement error, DSA was applied on one error-contaminated measure without measurement error correction (naive approach)). Average values computed over 200 simulation runs are displayed. Since FDP cannot be defined when there is a single exposure in the study and it is a true predictor (which means there is no exposure that is not a true predictor), we assumed as a baseline value, a theoretical FDP (in the absence of measurement error) of 5%.

DSA: Deletion/substitution/addition algorithm; FDP: False discovery proportion.

1. Mitigating the impact of measurement error in exposome studies

Regression calibration allowed some improvement over the naive approach ignoring measurement error when repeated biospecimens were available, and a good recovery of the impact of exposure measurement error when five repeated biospecimens were available per exposure and subject; there was little gain in further increasing the number of biospecimens per subject. When no repeated biospecimens are available, we cannot strongly recommend using RC, for the method did not always provide a clear improvement over the naive (uncorrected) approach; moreover "external" ICCs (i.e., stemming from another population) are needed in this case, and ICCs are not always transposable from one population to another (Vernet et al., 2019).

Our study can also be seen as a generalization in the exposome setting of the within-subject biospecimens pooling approach, which had previously been validated in a single-exposure context (Perrier et al., 2016; Vernet et al., 2019). We assumed that no source of error was introduced when pooling biospecimens. The performances of pooling were marginally lower than those of RC in our simulations. However, two things should be kept in mind: first, estimation bias can be improved by applying *a posteriori disattenuation*, which is efficient when an unbiased estimate of the ICC is available (preferably from a subgroup of the study population) (Perrier et al., 2016; Vernet et al., 2019). Second, RC implies to assay exposures in all collected biospecimens, as opposed to one per subject in the pooling approach. Consequently, biospecimens pooling can be achieved at a much lower cost, and allows reducing the issue related to limits of detection (LOD), as pooling limits the proportion of samples below the LOD (Schisterman and Vexler, 2008; Vernet et al., 2019).

The number of biospecimens collected per subject can be reduced without affecting significantly the methods' efficiency, preferably by limiting the assessment of repeated

biospecimens to exposures with the lowest ICCs (with 0.6 appearing as a relevant threshold); but also by assessing repeated samples in a subgroup of the study population (with 40% of randomly chosen subjects appearing as a minimal threshold in our setting). In practice, one should collect repeated urinary biospecimens (urine being the matrix from which the compounds with high within-subject variability are generally assessed) but could afford to collect only one or a couple of blood samples per subject, compounds assessed in blood generally having higher ICCs. It is not possible to suggest an optimum number of biospecimens to be collected per subject, as this depends on several factors, including the ICC and likely the correlation with other exposures. Yet, in a single exposure context, about 18-35 biospecimens per subject collected in the relevant time window were required to decrease bias in the dose-response slope down to 10% (Perrier et al., 2016; Vernet et al., 2019).

2. Study limitations

We considered a simple simulation design; specifically, we did not assume the existence of confounders (although these could be incorporated using our simulation code), of LODs (the performance gain of repeated biospecimen sampling might actually be larger if the LODs are high for some exposures, because collecting more biospecimens is generally a good strategy to limit issues related to LODs (Mumford et al., 2006)), nor of measurement error affecting the health outcome. We considered classical-type error only, which is typically what can be expected for biomarker-based exposures and which most existing statistical methods for measurement error are designed for. Yet departure from this type of error may be observed for exposures assessed by other means, such as atmospheric pollutants and meteorological conditions; and correlation in the errors across exposures may also exist (which RC can account for if the corresponding information is provided (Carroll et al., 2006)). Finally, we only considered linear effects of exposures on the health outcome; non-linear effects are likely to make the identification of true predictors even more challenging.

Regarding statistical approaches, we only investigated measurement error correction approaches that transformed the exposure variables, allowing flexibility in the algorithm to assess the link between exposures and the health parameter. Although we did not quantify it, we have no reason to believe that the impact of measurement error or of measurement error correction techniques would be different if another variable selection model was used at the exposome-health step. Several algorithms that jointly correct for measurement error and perform variable selection were recently developed (Liang and Li, 2009; Ma and Li, 2010; Sørensen et al., 2012; Wang et al., 2012), and can in principle be used in exposome studies. In practice, they are complex to implement, usually not available in standard statistical software and sometimes not applicable to all types of regression models or to all settings (Guangning, 2014).

Regarding our first step, instead of using RC for correcting for measurement error, moment reconstruction and multiple imputation may be considered. In the context of regression models without variable selection, a simulation study showed that RC carried efficiency gains that were sometimes dramatic over these two substitution methods (Freedman et al., 2008). Regarding the way RC was applied, here we have conducted RC correction only once for each individual scenario, including all available exposures measures in the model. Yet, unlike the pooling approach, RC is a multivariate model-based correction whose results depend on the set of exposures that are included in the model. Hence, ideally, RC correction should be refitted at each step of the DSA procedure with the updated set of retained exposures. This could improve the method performances, but appears cumbersome in practice. One alternative would be to correct each exposure independently. In balanced design scenarios without adjustment factors, this procedure is equivalent to the pooling approach when combined with a variable selection procedure, as the resulting exposure proxies are a rescaled version of the pooled biospecimens estimate, as we show in Appendix J. Finally, in the subsequent

exposure-health step, model's parameters standard errors are biased for both the RC and the pooling approaches (they do not account for the fact that exposures are measured with error (Armstrong and Basagaña, 2015; Spiegelman et al., 2001)). Proper standard error estimates may be obtained by bootstrap. Yet, sensitivity, FDP and estimation bias performances obtained with our DSA approach are not affected by estimation bias of standard errors, as this variable selection method does not depend on the coefficients standard error, but relies on the model fit.

3. Conclusions and possible strategy for future exposome studies

As demonstrated here and in previous studies, the factors that negatively affect the performances of statistical methods in studies linking exposome and health are notably (i) a large number of exposures ; (ii) the existence of (even moderate) correlation between exposures; and (iii) differential measurement error (Figure 5). Schematically, an exposome study that considers both persistent and non-persistent biomarkers risks being short-sighted (underpowered) for the least persistent exposures if no specific efforts to correct the related unbalanced power is made. Collecting repeated biospecimens and using within-subject pooling or regression calibration on these repeated biospecimens allows improving performances up to a large extent when multiple biospecimens are collected per exposure and subject. If assay cost is an issue, then the within-subject biospecimens pooling approach, which allows improving performances without increasing assay cost (but with an increase in biospecimens collection and handling costs compared to the approach with one biospecimen per subject) should be preferred; limiting repeated samples to the least persistent exposures is a further option to limit costs.

In conclusion, measurement error issues are compounded in exposome studies compared to smaller dimension exposure-health studies. Exposome research has the more or less overtly

stated aim to hierarchize exposures in terms of strength of their association with health; as we illustrated here (see in particular Figure 2), this aim is unlikely to be achieved without implementing, from the step of study design, specific measurement error correction approaches such as those relying on the collection of repeated biospecimens in each subject.

Acknowledgements: ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya. We thank Dr. Claire Philippat for her fruitful comments on the study design and statistical methods assessment, and Dr. Florent Chuffart for his help using the Ciment infrastructure.

Conflict of interests: The authors declare no conflict of interests.

REFERENCES

Agier, L., Basagaña, X., Hernandez-Ferrer, C., Maitre, L., Tamayo Uria, I., Urquiza, J., Andrusaityte, S., Casas, M., de Castro, M., Cequier, E., Chatzi, L., Donaire-Gonzalez, D., Giorgis-Allemand, L., Gonzalez, J.R., Grazuleviciene, R., Gützkow, K.B., Haug, L.S., Sakhi, A.K., McEachan, R.R.C., Meltzer, H.M., Nieuwenhuijsen, M., Robinson, O., Roumeliotaki, T., Sunyer, J., Thomsen, C., Vafeiadi, M., Valentin, A., West, J., Wright, J., Siroux, V., Vrijheid, M., Slama, R., 2020. Association between the pregnancy exposome and fetal growth (in press). Int. J. Epidemiol.

Agier, L., Basagaña, X., Maitre, L., Granum, B., Bird, P., Casas, M., Oftedal, B., Wright, J., Andrusaityte, S., de Castro, M., Cequier, E., Chatzi, L., Donaire-Gonzales, D., Grazuleviciene, R., Haug, L.S., Sakhi, A.K., Leventakou, V., McEachan, R., Nieuwenhuijsen, M., Petraviciene, I., Robinson, O., Roumeliotaki, T., Sunyer, J., Tamayo, I., Thomsen, C., Urquiza, J., Valentin, A., Slama, R., Vrijheid, M., Siroux, V., 2019. Early life exposome and lung function in children from the HELIX cohort. Lancet Planet. Heal. 3, e81_e92. https://doi.org/10.1016/S2542-5196(19)30010-5

Agier, L., Portengen, L., Chadeau-Hyam, M., Basagaña, X., Giorgis-Allemand, L., Siroux, V., Robinson, O., Vlaanderen, J., González, J.R., Nieuwenhuijsen, M.J., Vineis, P., Vrijheid, M., Slama, R., Vermeulen, R., 2016. A Systematic Comparison of Linear Regression-Based Statistical Methods to Assess Exposome-Health Associations. Environ. Health Perspect. 124, 1848–1856. https://doi.org/10.1289/EHP172

Armstrong, B., Basagaña, X., 2015. Exposure Measurement Error: Consequences and Design Issues, in: Exposure Assessment in Environmental Epidemiology. Oxford University Press. https://doi.org/10.1093/med/9780199378784.003.0010

Armstrong, B.G., 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup. Environ. Med. 55, 651–656. https://doi.org/10.1136/oem.55.10.651

Brakenhoff, T.B., Mitroiu, M., Keogh, R.H., Moons, K.G.M., Groenwold, R.H.H., van Smeden, M., 2018a. Measurement error is often neglected in medical literature: a systematic review. J. Clin. Epidemiol. 98, 89–97. https://doi.org/10.1016/j.jclinepi.2018.02.023

Brakenhoff, T.B., van Smeden, M., Visseren, F.L.J., Groenwold, R.H.H., 2018b. Random measurement error: Why worry? An example of cardiovascular risk factors. PLoS One 13, e0192298. https://doi.org/10.1371/journal.pone.0192298

Brown, B., Weaver, T., Wolfson, J., 2019. MEBoost: Variable selection in the presence of measurement error. Stat. Med. 38, 2705–2718. https://doi.org/10.1002/sim.8130

Buzas, J.S., Stefanski, L.A., Tosteson, T.D., 2005. Measurement Error, in: Handbook of Epidemiology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 729–765. https://doi.org/10.1007/978-3-540-26577-1_19

Calafat, A.M., Ye, X., Silva, M.J., Kuklenyik, Z., Needham, L.L., Kolossa, M., Tuomisto, J., Astrup Jensen, A., 2006. Human exposure assessment to environmental chemicals using biomonitoring, in: International Journal of Andrology. pp. 166–171. https://doi.org/10.1111/j.1365-2605.2005.00570.x

Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., 2006. Measurement Error in Nonlinear Models: A Modern Perspective, CRC Press. ed. https://doi.org/10.1111/j.1467-985X.2007.00528_4.x

Casas, M., Basagaña, X., Sakhi, A.K., Haug, L.S., Philippat, C., Granum, B., Manzano-Salgado, C.B., Brochot, C., Zeman, F., de Bont, J., Andrusaityte, S., Chatzi, L., Donaire-Gonzalez, D., Giorgis-Allemand, L., Gonzalez, J.R., Gracia-Lavedan, E.,

Grazuleviciene, R., Kampouri, M., Lyon-Caen, S., Pañella, P., Petraviciene, I., Robinson, O., Urquiza, J., Vafeiadi, M., Vernet, C., Waiblinger, D., Wright, J., Thomsen, C., Slama, R., Vrijheid, M., 2018. Variability of urinary concentrations of non-persistent chemicals in pregnant women and school-aged children. Environ. Int. 121, 561–573. https://doi.org/10.1016/j.envint.2018.09.046

de Klerk, N.H., English, D.R., Armstrong, B.K., 1989. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. Int. J. Epidemiol. 18, 705–712. https://doi.org/10.1093/ije/18.3.705

Donoho, D., Jin, J., 2008. Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. Proc. Natl. Acad. Sci. U. S. A. 105, 14790–14795. https://doi.org/10.1073/pnas.0807471105

Freedman, L.S., Midthune, D., Carroll, R.J., Kipnis, V., 2008. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. Stat. Med. 27, 5195–216. https://doi.org/10.1002/sim.3361

Guangning, X., 2014. Variable Selection in Measurement Error Models. https://doi.org/10.1017/CBO9781107415324.004

Guxens, M., Ballester, F., Espada, M., Fernández, M., Grimalt, J., Ibarluzea, J., Olea, N., Rebagliato, M., Tardón, A., Torrent, M., Vioque, J., Vrijheid, M., Sunyer, J., 2012. Cohort Profile: The INMA –INfancia y Medio Ambiente–(Environment and Childhood) Project. Int. J. Epidemiol. 930–940. https://doi.org/10.1093/ije/dyr054

Jurek, A.M., Greenland, S., Maldonado, G., Church, T.R., 2005. Proper interpretation of non-differential misclassification effects: expectations vs observations. Int. J. Epidemiol. 34, 680–687. https://doi.org/10.1093/ije/dyi060

LaKind, J.S., Idri, F., Naiman, D.Q., Verner, M.A., 2019. Biomonitoring and Nonpersistent Chemicals-Understanding and Addressing Variability and Exposure Misclassification. Curr. Environ. Heal. reports. https://doi.org/10.1007/s40572-019-0227-2

Liang, H., Li, R., 2009. Variable Selection for Partially Linear Models with Measurement Errors. J. Am. Stat. Assoc. 104, 234–248. https://doi.org/10.1198/jasa.2009.0127

Lyon-Caen, S., Siroux, V., Lepeule, J., Lorimier, P., Hainaut, P., Mossuz, P., Quentin, J., Supernant, K., Meary, D., Chaperot, L., Bayat, S., Cassee, F., Valentino, S., Couturier-Tarrade, A., Rousseau-Ralliard, D., Chavatte-Palmer, P., Philippat, C., Pin, I., Slama, R., 2019. Deciphering the impact of early-life exposures to highly variable environmental factors on foetal and child health: Design of SEPAGES couple-child cohort. Int. J. Environ. Res. Public Health 16, 3888. https://doi.org/10.3390/ijerph16203888

Ma, Y., Li, R., 2010. Variable Selection in Measurement Error Models. Bernoulli (Andover). 16, 274–300. https://doi.org/10.3150/09-BEJ205

Mumford, S.L., Schisterman, E.F., Vexler, A., Liu, A., 2006. Pooling biospecimens and limits of detection: effects on ROC curve analysis. Biostatistics 7, 585–598. https://doi.org/10.1093/biostatistics/kxj027

Perrier, F., Giorgis-Allemand, L., Slama, R., Philippat, C., 2016. Within-subject Pooling of Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies. Epidemiology 27, 378–388. https://doi.org/10.1097/EDE.0000000000000460

Preau, J.L., Wong, L.Y., Silva, M.J., Needham, L.L., Calafat, A.M., 2010. Variability over 1 week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among eight adults: An observational study. Environ. Health Perspect. 118, 1748–1754. https://doi.org/10.1289/ehp.1002231

Schisterman, E.F., Albert, P.S., 2012. The biomarker revolution. Stat. Med. 31, 2513–5. https://doi.org/10.1002/sim.5499

Schisterman, E.F., Vexler, A., 2008. To pool or not to pool, from whether to when:

Applications of pooling to biospecimens subject to a limit of detection. Paediatr. Perinat. Epidemiol. 22, 486–496. https://doi.org/10.1111/j.1365-3016.2008.00956.x
Sinisi, S.E., van der Laan, M.J., 2004. Deletion/Substitution/Addition Algorithm in Learning with Applications in Genomics. Stat. Appl. Genet. Mol. Biol. 3, 1–38. https://doi.org/10.2202/1544-6115.1069
Slama, R., Vrijheid, M., 2015. Some Challenges of Studies Aiming to Relate the Exposome to Human Health. Occup. Environ. Med. 0.
Sørensen, Ø., Frigessi, A., Thoresen, M., 2012. Measurement Error in Lasso: Impact and Correction 1–64. https://doi.org/10.5705/ss.2013.180
Spiegelman, D., Carroll, R.J., Kipnis, V., 2001. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. Stat. Med. 20, 139–160. https://doi.org/10.1002/1097-0258(20010115)20:1<139::AID-SIM644>3.0.CO;2-K
Tamayo-Uria, I., Maitre, L., Thomsen, C., Nieuwenhuijsen, M.J., Chatzi, L., Siroux, V., Aasvang, G.M., Agier, L., Andrusaityte, S., Casas, M., de Castro, M., Dedele, A., Haug, L.S., Heude, B., Grazuleviciene, R., Gutzkow, K.B., Krog, N.H., Mason, D., McEachan, R.R.C., Meltzer, H.M., Petraviciene, I., Robinson, O., Roumeliotaki, T., Sakhi, A.K., Urquiza, J., Vafeiadi, M., Waiblinger, D., Warembourg, C., Wright, J., Slama, R., Vrijheid, M., Basagaña, X., 2019. The early-life exposome: Description and patterns in six European countries. Environ. Int. 123, 189–200. https://doi.org/10.1016/J.ENVINT.2018.11.067
Vasquez, M.M., Hu, C., Roe, D.J., Halonen, M., Guerra, S., 2019. Measurement error correction in the least absolute shrinkage and selection operator model when validation data are available. Stat. Methods Med. Res. 28, 670–680. https://doi.org/10.1177/0962280217734241
Vernet, C., Philippat, C., Agier, L., Calafat, A.M., Ye, X., Lyon-Caen, S., Hainaut, P., Siroux, V., Schisterman, E.F., Slama, R., 2019. An empirical validation of the within-subject biospecimens pooling approach to minimize exposure misclassification in biomarker-based studies. Epidemiology. https://doi.org/10.1097/ede.0000000000001056
Vrijheid, M., Slama, R., Robinson, O., Chatzi, L., Coen, M., van den Hazel, P., Thomsen, C., Wright, J., Athersuch, T.J., Avellana, N., Basagaña, X., Brochot, C., Bucchini, L., Bustamante, M., Carracedo, A., Casas, M., Estivill, X., Fairley, L., van Gent, D., Gonzalez, J.R., Granum, B., Gražulevičienė, R., Gutzkow, K.B., Julvez, J., Keun, H.C., Kogevinas, M., McEachan, R.R.C., Meltzer, H.M., Sabidó, E., Schwarze, P.E., Siroux, V., Sunyer, J., Want, E.J., Zeman, F., Nieuwenhuijsen, M.J., 2014. The Human Early-Life Exposome (HELIX): Project Rationale and Design. Environ. Health Perspect. 122, 535–544. https://doi.org/10.1289/ehp.1307204
Wang, H.J., Stefanski, L.A., Zhu, Z., 2012. Corrected-loss estimation for quantile regression with covariate measurement errors. Biometrika 99, 405–421. https://doi.org/10.1093/biomet/ass005
Wielgomas, B., 2013. Variability of urinary excretion of pyrethroid metabolites in seven persons over seven consecutive days-Implications for observational studies. Toxicol. Lett. 221, 15–22. https://doi.org/10.1016/j.toxlet.2013.05.009
Wild, C.P., 2012. The exposome: from concept to utility. Int. J. Epidemiol. 41, 24–32. https://doi.org/10.1093/ije/dyr236
Ye, X., Wong, L.-Y., Bishop, A.M., Calafat, A.M., 2011. Variability of urinary concentrations of bisphenol A in spot samples, first morning voids, and 24-hour collections. Environ. Health Perspect. 119, 983–8. https://doi.org/10.1289/ehp.1002701