Supplemental material

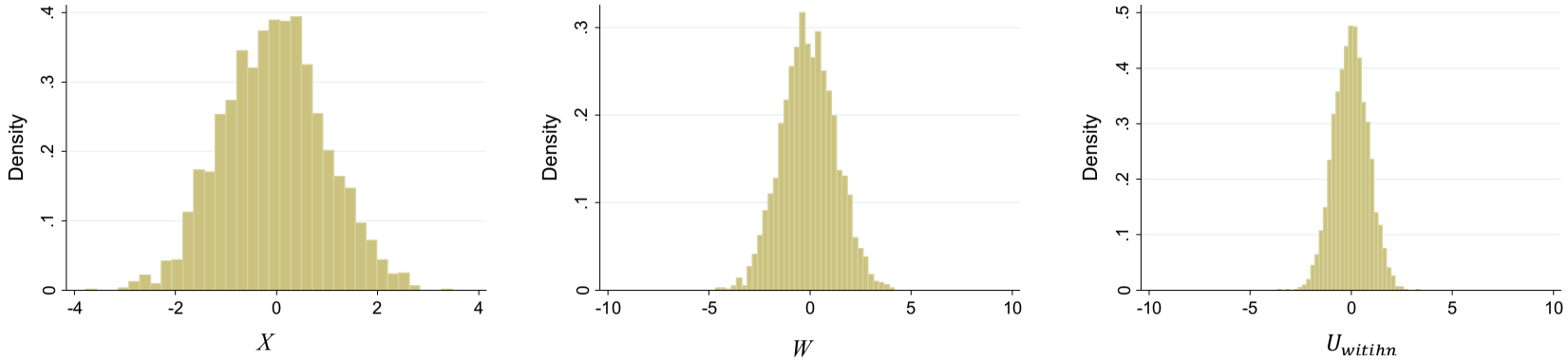
Within-subject Pooling of Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies

Flavie Perrier, Lise Giorgis-Allemand, Remy Slama, Claire Philippat

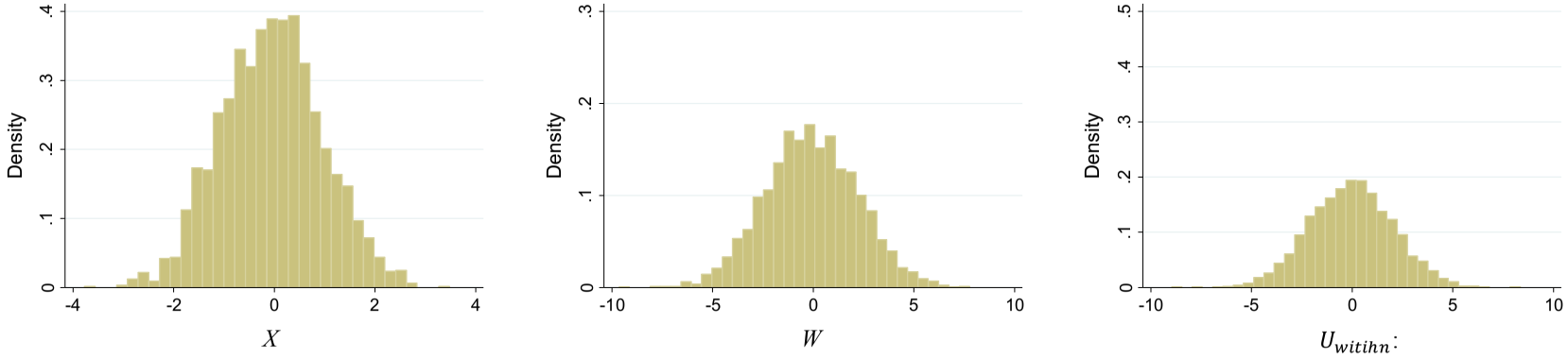
Table of contents

eFigure 1: Distribution of the simulated exposure (X, W) and the measurement error (U_{within}) in one of our simulated studies for chemicals with an ICC of 0.6 (A) and 0.2 (B) (lack of between-assay error).

A) ICC of 0.6



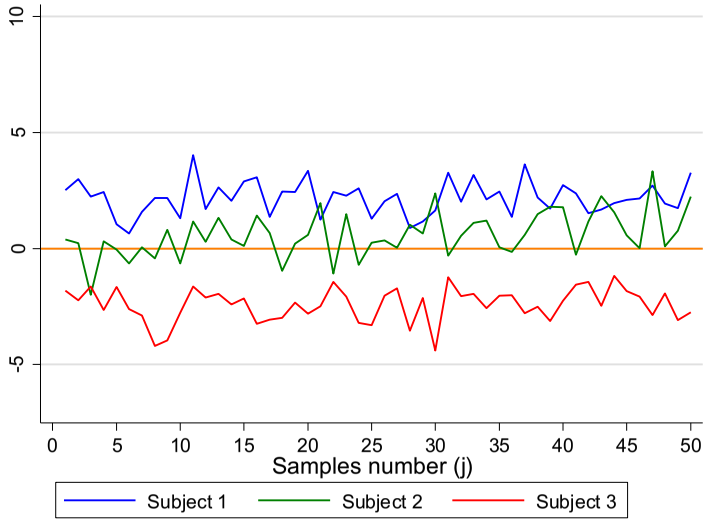
B) ICC of 0.2



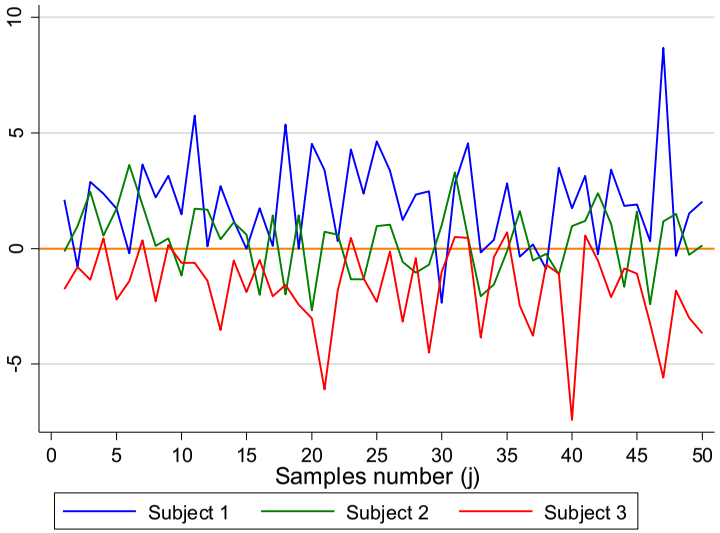
Abbreviations: U_{witihn}: measurement error, : W : error-prone variable, X: real exposure measured without error

eFigure 2: Normalized urine concentrations of chemical A (ICC, 0.6) and chemical B (ICC, 0.2), measured with error, for 3 participants of one of our simulated studies (lack of between-assay error variance).

A) Compound A with an ICC of 0.6



B) Compound B with an ICC of 0.2



The horizontal line represents the real unobserved average of urinary concentrations of the considered chemical over a toxicologically relevant exposure window, measured without error and centered around 0.

eTable 1: Effect of the between-assay error on the effect estimates and statistical power of studies aiming at characterizing the associations between biomarker-based exposure and a continuous health outcome (1000 simulations of studies with 3000 subjects each; real effect β_1 = -100 g). ICC decreased with increasing between-assay error.

				Chemical A												Chemical B											
σ_{Usssay}^2	Number of specimens per subject	Real effect[a]	ICC[a]	Within-subject pooling			Within-subject pooling + dissatenuation			Simex			Regression calibration			ICC[a]	Within-subject pooling			Within-subject pooling + dissatenuation			Simex			Regression calibration		
				Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]		Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]
	1	-100	0.60	-60[e]	0.71	40	-100	0.71	0							0.20	-20[e]	0.32	80	-100	0.32	0						
	2	-100	0.60	-75	0.82	25	-100	0.82	0	-95	0.85	5	-100	0.86	0	0.20	-34	0.50	66	-101	0.50	1	-54	0.52	45	-101	0.53	2
0	**5**	-100	0.60	-89	0.88	11	-100	0.88	0	-100	0.89	0	-101	0.91	1	0.20	-56	0.71	43	-101	0.71	1	-82	0.72	18	-102	0.74	2
	10	-99	0.60	-93	0.90	6	-100	0.90	1	-100	0.91	0	-100	0.91	0	0.20	-72	0.80	28	-100	0.8	1	-93	0.82	6	-100	0.83	1
	50	-99	0.60	-97	0.90	2	-98	0.90	1	-98	0.92	0	-98	0.92	0	0.20	-91	0.88	8	-98	0.88	1	-98	0.90	1	-98	0.90	0
	1	-101	0.60	-60[e]	0.72	40	-100	0.72	1							0.20	-20[e]	0.33	80	-101	0.33	0						
	2	-99	0.60	-74	0.81	26	-98	0.81	1	-94	0.83	6	-99	0.84	0	0.20	-33	0.47	67	-98	0.47	1	-53	0.51	47	-99	0.51	0
0.01	**5**	-101	0.60	-88	0.88	13	-100	0.88	1	-100	0.89	1	-101	0.9	0	0.20	-56	0.70	45	-100	0.70	1	-81	0.73	20	-101	0.74	0
	10	-100	0.60	-93	0.89	7	-99	0.89	1	-100	0.91	0	-100	0.92	0	0.20	-71	0.79	29	-99	0.79	1	-93	0.80	7	-100	0.83	0
	50	-99	0.60	-97	0.90	2	-98	0.90	1	-99	0.92	0	-99	0.92	0	0.20	-91	0.88	8	-98	0.88	1	-99	0.89	0	-99	0.90	0
	1	-101	0.57	-57	0.71	43	-95	0.71	6							0.196	-20[e]	0.32	80	-99	0.32	2						
	2	-99	0.57	-69	0.79	31	-92	0.79	7	-93	0.82	7	-99	0.83	0	0.196	-32	0.46	68	-95	0.46	4	-52	0.51	47	-99	0.50	0
0.1	**5**	-101	0.57	-82	0.86	19	-93	0.86	8	-100	0.89	1	-101	0.89	0	0.196	-53	0.68	47	-95	0.68	6	-80	0.72	20	-101	0.74	0
	10	-100	0.57	-86	0.87	14	-92	0.87	8	-100	0.91	0	-100	0.91	0	0.196	-67	0.76	33	-94	0.76	6	-93	0.80	7	-100	0.83	0
	50	-99	0.57	-89	0.88	10	-90	0.88	9	-99	0.92	0	-99	0.92	0	0.196	-84	0.85	15	-91	0.85	8	-99	0.90	0	-99	0.90	0
	1	-101	0.46	-47	0.62	54	-78	0.62	23							0.18	-18[e]	0.29	82	-92	0.29	9						
	2	-99	0.46	-54	0.68	46	-72	0.69	27	-86	0.78	13	-99	0.79	0	0.18	-28	0.42	72	-84	0.42	15	-50	0.49	50	-99	0.48	0
0.5	**5**	-101	0.46	-62	0.75	39	-70	0.75	31	-98	0.87	3	-101	0.87	0	0.18	-44	0.60	57	-79	0.6	22	-78	0.71	23	-101	0.71	0
	10	-100	0.46	-64	0.77	36	-69	0.77	31	-100	0.90	0	-100	0.90	0	0.18	-53	0.67	47	-74	0.67	26	-92	0.79	9	-100	0.82	0
	50	-99	0.46	-66	0.76	34	-66	0.76	33	-99	0.92	0	-99	0.92	0	0.18	-63	0.74	37	-68	0.74	32	-99	0.90	0	-99	0.90	0

[a] ICC decreased with increasing between-assay error variance.

[b] Mean of 1000 effect estimates. SIMEX and regression calibration cannot be used if only one biospecimen per subject is available.

[c] Statistical 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_{ij}) and the continuous outcome was below 0.05.
[d] Difference between the real effect and the effect estimate, divided by the real effect.
[e] Corresponds to a situation without pooling.

Note: The between assay error ($U_{ij,assay}$) was randomly drawn from a distribution with mean 0 and variance ${\sigma_{Uassay}}^2$. With regard to the relatively low coefficients of variation reported for short half-life chemicals (for example, coefficients of variation of about 60 replicates in a period of 9 months were between 3% and 10% for phenol biomarkers[1]), we expected the between-assay error to be of low amplitude compared to the total variability. We however used a relatively large set of values for the between-assay variance (${\sigma_{Uassay}}^2$), ranging from 0.01, 0.1 to 0.5 (half of the variance of X_i). We hypothesize that each measured concentration (W_{ij}) in the j^{th} biospecimen of subject i was affected by the between-assay error, in addition to the error related to the within-subject error, such that:

$$W_{ij} = X_i + U_{ij,within} + U_{ij,assay} \qquad i = 1, \dots, n \; ; j = 1, \dots, k$$

The concentration measured in the pool of k biospecimens of equal volume from subject i was simulated as follows:

$$W_{ik,pool} = X_i + \frac{\sum_{j=1}^{k} U_{ij,indiv}}{k} + U_{ik,assay} \qquad i = 1, \dots, n; \; k = 1, \dots, 50$$

(This formula assumes that no chemical reaction takes place in the biospecimens before or after pooling).

With this approach, ICC decreased with increasing between-assay error, which makes it difficult to differentiate between the effect on bias due to the within-subject error ($U_{ij,within}$) and that due to the between-assay error. An alternative approach in which the ICC was kept constant whatever the value of the between-assay variance is presented in eTable 2.

eTable 2: Effect of the between-assay error on the effect estimates and statistical power of studies aiming at characterizing the associations between biomarker-based exposure and a continuous health outcome (1000 simulations of studies with 3000 subjects each; real effect β_1 = -100 g). ICC was kept constant whatever the value of the between-assay error.

			Chemical A, ICC of 0.6[a]												Chemical B, ICC of 0.2[a]											
			Within-subject pooling			Within-subject pooling + dissatenuation			Simex			Regression calibration			Within-subject pooling			Within-subject pooling + dissatenuation			Simex			Regression calibration		
σ_{Usssay}^2	Number of specimens per subject	Real effect[b]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]	Effect estimate[b]	Power[c]	Bias (%)[d]
	1	-100	-60[e]	0.71	40	-100[e]	0.71	0							-20[e]	0.32	80	-100[e]	0.32	0						
	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	-54	0.52	45	-101	0.53	2
0	**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	-82	0.72	18	-102	0.74	2
	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	-93	0.82	6	-100	0.83	1
	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
	1	-100[e]	-60[e]	0.73	40	-100[e]	0.73	0							-20[e]	0.31	80	-100[e]	0.31	0						
	2	-101	-75	0.82	25	-101	0.82	0	-96	0.84	5	-101	0.85	0	-34	0.50	67	-101	0.50	0	-55	0.53	46	-102	0.53	1
0.01	**5**	-100	-87	0.88	13	-99	0.88	1	-99	0.90	1	-100	0.90	0	-55	0.68	45	-99	0.68	1	-80	0.72	20	-99	0.73	1
	10	-99	-92	0.89	7	-98	0.89	1	-99	0.91	0	-99	0.91	0	-70	0.80	29	-99	0.80	0	-92	0.82	7	-99	0.83	0
	50	-101	-99	0.91	2	-100	0.91	1	-101	0.93	0	-101	0.92	0	-93	0.89	8	-100	0.89	1	-101	0.90	0	-101	0.91	0
	1	-100[e]	-60[e]	0.74	40	-100[e]	0.74	0							-20[e]	0.31	80	-100[e]	0.32	0						
	2	-101	-73	0.80	27	-98	0.80	3	-96	0.84	5	-101	0.85	0	-33	0.50	67	-100	0.51	1	-55	0.53	46	-102	0.53	1
0.1	**5**	-100	-82	0.86	18	-93	0.86	7	-99	0.90	1	-100	0.90	0	-53	0.66	47	-95	0.66	5	-80	0.71	20	-99	0.72	1
	10	-99	-86	0.87	13	-92	0.87	7	-99	0.91	0	-99	0.91	0	-67	0.79	33	-93	0.79	6	-92	0.81	7	-99	0.82	0
	50	-101	-91	0.88	10	-92	0.88	9	-101	0.93	0	-101	0.92	0	-86	0.86	15	-93	0.86	8	-101	0.90	0	-101	0.91	0
	1	-100[e]	-60[e]	0.74	40	-100[e]	0.74	0							-20[e]	0.32	80	-100[e]	0.32	0						
	2	-101	-64	0.75	36	-86	0.75	15	-96	0.85	5	-101	0.87	0	-31	0.48	69	-94	0.48	7	-55	0.54	46	-102	0.53	1
0.5	**5**	-100	-65	0.76	35	-73	0.76	27	-99	0.90	1	-100	0.90	0	-45	0.59	55	-81	0.59	19	-80	0.72	20	-99	0.72	1
	10	-99	-66	0.77	34	-70	0.77	29	-99	0.91	0	-99	0.91	0	-54	0.68	46	-76	0.68	24	-92	0.82	7	-99	0.83	0
	50	-101	-67	0.77	33	-68	0.77	32	-101	0.93	0	-101	0.92	0	-65	0.76	36	-70	0.76	31	-101	0.90	0	-101	0.90	0

[a] ICC was constant whatever the value of the between-assay error variance.

[b] Mean of 1000 effect estimates. SIMEX and regression calibration cannot be used if only one biospecimen per subject is available.

[c] Statistical 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_{ij}) and the continuous outcome was below 0.05.

[d] Difference between the real effect and the effect estimate, divided by the real effect.

[e] Corresponds to a situation without pooling.

Note: The simulation of the impact of between-assay error slightly differed compared to eTable 1. In eTable 1, the between assay error was added to the $U_{ij,within}$, while in eTable 6, we broke up the within-subject error $U_{ij,within}$ in two components: 1) the part of the error related to the individual ($U_{ij,indiv}$), 2) the between-assay error ($U_{ij,assay}$) due to the measurement such that:

$$U_{ij,within} = U_{ij,indiv} + U_{ij,assay} \qquad i = 1, \dots, n \; ; \; j = 1, \dots, k \qquad (2)$$

With this approach, any increase in between-assay error is counterbalanced by a decrease in the error related to the individual ($U_{ij,indiv}$). Thus, the ICC is maintained constant (which was not the case in eTable 1). This might not be realistic but allows discriminating an effect of the between-assay error from an effect of the within-subject variability; it allows to discriminate, for a given ICC, which approach (pooling, SIMEX, regression calibration) is the most efficient to correct bias, according to whether the error predominantly comes from the individual or the assay.

eTable 3: Binary outcome, ICC of 0.6 - Effect estimates, Odds-Ratios and statistical power of the simulated studies aiming at characterizing the association between biomarker-based exposure to chemical A (ICC, 0.6) and a binary health outcome, according to the number of biospecimens collected per subject and the approach used to limit exposure misclassification (1000 simulations of studies with 3000 subjects each; real effect $\beta_1 = 0.262$ corresponding to an OR of 1.30, assuming lack of between-assay error)

Number of biospecimens per subject	Real effect[b]	Within-subject pooling					Within-subject pooling + a posteriori dissatenuation					SIMEX					Regression calibration				
		OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)
1	0.264	1.17	0.159	[0.03 ; 0.29]	0.69	40	1.31	0.264	[0.05 ; 0.49]	0.69	0										
2	0.266	1.22	0.198	[0.05 ; 0.34]	0.77	25	1.31	0.265	[0.06 ; 0.45]	0.77	1	1.29	0.251	[0.06 ; 0.43]	0.77	5	1.31	0.265	[0.06 ; 0.46]	0.79	0
3	0.262	1.24	0.215	[0.07 ; 0.35]	0.82	18	1.31	0.263	[0.08 ; 0.43]	0.82	0	1.30	0.258	[0.08 ; 0.43]	0.82	2	1.31	0.263	[0.08 ; 0.43]	0.83	0
4	0.260	1.25	0.221	[0.06 ; 0.37]	0.81	15	1.30	0.258	[0.07 ; 0.44]	0.81	1	1.30	0.255	[0.07 ; 0.43]	0.82	2	1.30	0.258	[0.07 ; 0.44]	0.82	1
5	0.265	1.27	0.234	[0.09 ; 0.39]	0.86	12	1.31	0.265	[0.10 ; 0.44]	0.86	0	1.31	0.263	[0.09 ; 0.44]	0.85	1	1.31	0.265	[0.10 ; 0.44]	0.87	0
6	0.259	1.27	0.233	[0.08 ; 0.39]	0.83	10	1.30	0.259	[0.08 ; 0.43]	0.83	0	1.30	0.258	[0.08 ; 0.42]	0.85	1	1.30	0.259	[0.08 ; 0.43]	0.85	0
7	0.261	1.27	0.238	[0.07 ; 0.39]	0.84	9	1.30	0.261	[0.08 ; 0.43]	0.84	0	1.30	0.260	[0.08 ; 0.43]	0.85	0	1.30	0.261	[0.08 ; 0.43]	0.86	0
8	0.259	1.28	0.240	[0.09 ; 0.40]	0.86	7	1.30	0.260	[0.10 ; 0.43]	0.86	0	1.30	0.259	[0.09 ; 0.43]	0.87	0	1.30	0.260	[0.10 ; 0.43]	0.87	0
9	0.262	1.28	0.244	[0.08 ; 0.40]	0.86	7	1.30	0.262	[0.09 ; 0.43]	0.86	0	1.30	0.262	[0.08 ; 0.43]	0.87	0	1.30	0.262	[0.09 ; 0.43]	0.88	0
10	0.261	1.28	0.245	[0.08 ; 0.41]	0.85	6	1.30	0.261	[0.08 ; 0.44]	0.85	0	1.30	0.261	[0.08 ; 0.44]	0.86	0	1.30	0.261	[0.08 ; 0.44]	0.87	0
12	0.264	1.29	0.250	[0.09 ; 0.41]	0.86	5	1.31	0.264	[0.09 ; 0.44]	0.86	0	1.31	0.264	[0.09 ; 0.44]	0.87	0	1.31	0.264	[0.09 ; 0.44]	0.88	0
15	0.262	1.29	0.251	[0.09 ; 0.41]	0.88	4	1.30	0.262	[0.09 ; 0.42]	0.88	0	1.30	0.262	[0.09 ; 0.43]	0.88	0	1.30	0.262	[0.09 ; 0.42]	0.89	0
18	0.265	1.30	0.256	[0.09 ; 0.43]	0.88	4	1.31	0.265	[0.09 ; 0.44]	0.88	0	1.31	0.265	[0.09 ; 0.45]	0.88	0	1.31	0.265	[0.09 ; 0.44]	0.89	0
20	0.267	1.30	0.258	[0.09 ; 0.42]	0.88	3	1.31	0.266	[0.10 ; 0.43]	0.88	0	1.31	0.266	[0.10 ; 0.43]	0.90	0	1.31	0.266	[0.10 ; 0.43]	0.89	0
25	0.266	1.30	0.259	[0.08 ; 0.42]	0.88	3	1.31	0.266	[0.09 ; 0.43]	0.88	0	1.31	0.266	[0.09 ; 0.44]	0.89	0	1.31	0.266	[0.09 ; 0.43]	0.88	0
30	0.260	1.29	0.254	[0.09 ; 0.42]	0.87	2	1.30	0.260	[0.09 ; 0.42]	0.87	0	1.30	0.260	[0.09 ; 0.42]	0.88	0	1.30	0.260	[0.09 ; 0.42]	0.88	0
35	0.260	1.30	0.255	[0.09 ; 0.42]	0.86	2	1.30	0.260	[0.09 ; 0.43]	0.86	0	1.30	0.260	[0.09 ; 0.43]	0.88	0	1.30	0.260	[0.09 ; 0.43]	0.87	0
40	0.266	1.30	0.262	[0.09 ; 0.42]	0.89	2	1.31	0.266	[0.09 ; 0.43]	0.89	0	1.31	0.267	[0.09 ; 0.43]	0.90	0	1.31	0.266	[0.09 ; 0.43]	0.90	0
50	0.264	1.30	0.261	[0.09 ; 0.43]	0.88	1	1.31	0.264	[0.09 ; 0.43]	0.88	0	1.31	0.264	[0.09 ; 0.43]	0.88	0	1.31	0.264	[0.09 ; 0.43]	0.89	0

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

[a] Mean of 1000 estimated odds ratio.

[b] Mean of 1000 effect estimates.

[c] Empirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1000 simulation runs.

[d] Statistical power, proportion of studies in which the p-value of the parameter characterizing the association between the error-prone variables (W_{ij}) and the binary outcome was below 0.05.

[e] Difference between the real effect and the effect estimate, divided by the real effect.

[f] Corresponds to a situation without pooling.

Note: The binary outcome (Y') was simulated as follows. First, $P(Y'_i = 1)$ was estimated according to a logistic regression model:

$$P(Y'_i = 1) = \frac{1}{1+\exp(-(\alpha' + \beta'_1 X_i))}$$

where α' = -2.944 (corresponding to a disease frequency of 5% in subjects with X = 0). The value of the parameter (β'_1) associated with real exposure X was 0.262, corresponding to an odds-ratio of 1.30 for each increase by 1 in exposure. For each subject, we then generated a random number between 0 and 1 drawn from a uniform distribution; we considered that the disease was present (Y'_i = 1) when the value of this random number was below $P(Y'_i = 1)$.

eTable 4: Binary outcome, ICC of 0.2 - Effect estimates, Odds-Ratios and statistical power of the simulated studies aiming at characterizing the association between a biomarker-based exposure to chemical B (ICC, 0.2) and a binary health outcome, according to the number of biospecimens collected per subject and the approach used to limit exposure misclassification (1000 simulations of studies with 3000 subjects each; real effect, $\beta_1 = 0.262$, corresponding to an OR of 1.30, assuming lack of between-assay error)

Number of biospecimens per subject	Real effect[b]	Within-subject pooling					Within-subject pooling + a posteriori dissatenuation					SIMEX					Regression calibration				
		OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)	OR[a]	Effect estimate[b]	95% CI[c]	Power[d]	Bias[e] (%)
1	0.264	1.06[f]	0.053	[-0.02 ; 0.13]	0.31	80	1.33[f]	0.265	[-0.11 ; 0.65]	0.31	0										
2	0.266	1.09	0.088	[-0.01 ; 0.19]	0.44	67	1.31	0.263	[-0.04 ; 0.56]	0.44	1	1.16	0.142	[-0.02 ; 0.31]	0.48	47	1.32	0.265	[-0.04 ; 0.58]	0.47	0
3	0.262	1.12	0.113	[0.00 ; 0.22]	0.56	57	1.31	0.264	[0.01 ; 0.50]	0.56	1	1.20	0.176	[0.01 ; 0.34]	0.57	33	1.31	0.265	[0.01 ; 0.50]	0.59	1
4	0.260	1.14	0.128	[0.01 ; 0.24]	0.58	51	1.3	0.255	[0.01 ; 0.48]	0.58	2	1.22	0.191	[0.01 ; 0.36]	0.60	26	1.30	0.256	[0.01 ; 0.48]	0.61	2
5	0.265	1.16	0.147	[0.02 ; 0.27]	0.64	45	1.31	0.264	[0.04 ; 0.49]	0.64	0	1.24	0.213	[0.03 ; 0.39]	0.66	20	1.31	0.264	[0.04 ; 0.49]	0.66	0
6	0.259	1.17	0.156	[0.02 ; 0.28]	0.68	40	1.30	0.259	[0.04 ; 0.47]	0.68	0	1.25	0.219	[0.04 ; 0.39]	0.70	15	1.30	0.259	[0.04 ; 0.47]	0.70	0
7	0.261	1.18	0.165	[0.04 ; 0.30]	0.71	37	1.30	0.260	[0.06 ; 0.46]	0.71	1	1.26	0.228	[0.05 ; 0.41]	0.71	13	1.30	0.260	[0.06 ; 0.46]	0.72	1
8	0.259	1.19	0.174	[0.05 ; 0.31]	0.73	33	1.30	0.261	[0.07 ; 0.47]	0.73	1	1.27	0.234	[0.07 ; 0.42]	0.73	10	1.30	0.261	[0.07 ; 0.47]	0.74	1
9	0.262	1.20	0.181	[0.03 ; 0.32]	0.76	31	1.31	0.262	[0.05 ; 0.46]	0.76	0	1.28	0.240	[0.05 ; 0.42]	0.76	9	1.31	0.262	[0.05 ; 0.46]	0.77	0
10	0.261	1.21	0.186	[0.04 ; 0.33]	0.76	29	1.30	0.261	[0.05 ; 0.46]	0.76	0	1.28	0.243	[0.04 ; 0.44]	0.78	7	1.30	0.261	[0.05 ; 0.47]	0.78	0
12	0.264	1.22	0.198	[0.06 ; 0.35]	0.78	25	1.31	0.264	[0.08 ; 0.46]	0.78	0	1.29	0.251	[0.07 ; 0.44]	0.77	5	1.31	0.264	[0.08 ; 0.46]	0.80	0
15	0.262	1.23	0.207	[0.06 ; 0.36]	0.80	21	1.31	0.262	[0.08 ; 0.45]	0.80	0	1.29	0.254	[0.07 ; 0.44]	0.81	3	1.31	0.262	[0.08 ; 0.45]	0.82	0
18	0.265	1.25	0.217	[0.06 ; 0.38]	0.83	18	1.31	0.265	[0.08 ; 0.46]	0.83	0	1.30	0.260	[0.08 ; 0.45]	0.84	2	1.31	0.265	[0.08 ; 0.46]	0.85	0
20	0.267	1.25	0.221	[0.07 ; 0.37]	0.84	17	1.31	0.266	[0.08 ; 0.45]	0.84	1	1.30	0.261	[0.08 ; 0.44]	0.84	2	1.31	0.266	[0.08 ; 0.45]	0.85	0
25	0.266	1.26	0.229	[0.07 ; 0.39]	0.83	14	1.31	0.266	[0.08 ; 0.45]	0.83	0	1.31	0.264	[0.08 ; 0.45]	0.85	1	1.31	0.266	[0.08 ; 0.45]	0.85	0
30	0.260	1.26	0.229	[0.07 ; 0.38]	0.84	12	1.30	0.260	[0.07 ; 0.43]	0.84	0	1.30	0.259	[0.07 ; 0.42]	0.85	0	1.30	0.260	[0.08 ; 0.43]	0.84	0
35	0.260	1.27	0.233	[0.07 ; 0.39]	0.83	11	1.30	0.259	[0.08 ; 0.44]	0.83	0	1.30	0.258	[0.08 ; 0.43]	0.84	1	1.30	0.259	[0.08 ; 0.44]	0.85	0
40	0.266	1.28	0.242	[0.08 ; 0.39]	0.87	9	1.31	0.266	[0.09 ; 0.43]	0.87	0	1.31	0.266	[0.09 ; 0.43]	0.87	0	1.31	0.266	[0.08 ; 0.43]	0.88	0
50	0.264	1.28	0.245	[0.09 ; 0.40]	0.85	7	1.31	0.264	[0.09 ; 0.43]	0.85	0	1.31	0.264	[0.09 ; 0.44]	0.86	0	1.31	0.264	[0.09 ; 0.43]	0.87	0

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

[a] Mean of 1000 estimated odds ratio.

[b] Mean of 1000 effect estimates.

[c] Empirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1000 simulation runs.

[d] Statistical power, proportion of studies in which the p-value of the parameter characterizing the association between the error-prone variables (W_{ij}) and the binary outcome was below 0.05.

[e] Difference between the real effect and the effect estimate, divided by the real effect.

[f] Corresponds to a situation without pooling.

Note: The binary outcome (Y') was simulated as explained in the note below eTable 3.

eTable 5: Effect estimates and type I error of studies aiming at characterizing the association between biomarker-based exposure to chemical A (ICC, 0.6) and B (ICC, 0.2) and a continuous outcome, according to the number of biospecimens collected per subject and the approach used to limit the effect of exposure misclassification (1000 simulations of studies with 3000 subjects each; real effect β_1 = 0 g, assuming lack of between-assay error).

Number of biospecimens per subject	Within-subject pooling			Within-subject pooling + dissatenuation			SIMEX			Regression calibration		
	Effect estimate[a]	95% CI[b]	Type I error[c]	Effect estimate[a]	95% CI[b]	Type I error[c]	Effect estimate[a]	95% CI[b]	Type I error[c]	Effect estimate[a]	95% CI[b]	Type I error[c]
Compound A, ICC of 0.6												
1	0 [d]	[-45; 45]	5%	0 [d]	[-76; 75]	5%						
2	0	[-52; 52]	5%	0	[-69; 70]	5%	0	[-66; 67]	7%	0	[-69; 70]	7%
5	0	[-56; 57]	6%	0	[-64; 64]	6%	-1	[-66; 63]	8%	0	[-64; 64]	7%
10	0	[-56; 56]	4%	0	[-60; 60]	4%	0	[-59; 61]	6%	0	[-60; 60]	6%
50	2	[-56; 59]	4%	2	[-56; 60]	4%	2	[-58; 58]	7%	2	[-56; 60]	7%
Compound B, ICC of 0.2												
1	0 [d]	[-25; 27]	5%	0 [d]	[-124; 136]	5%						
2	0	[-34; 35]	5%	-1	[-103; 105]	5%	0	[-56; 58]	6%	-1	[-99; 110]	6%
5	-1	[-49; 42]	5%	-1	[-88; 76]	5%	-1	[-72; 62]	7%	-2	[-88; 76]	7%
10	0	[-47; 48]	4%	0	[-66; 67]	4%	0	[-61; 65]	5%	0	[-66; 67]	5%
50	2	[-53; 58]	5%	2	[-57; 63]	5%	2	[-57; 64]	7%	2	[-57; 63]	7%

[a] Mean of 1000 effect estimates.

[b] Empirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1000 simulation runs.

[c] Type I error (false positive association), proportion of studies in which the p-value of the parameter characterizing the association between the error-prone variables and health outcome was below 0.05.

[d] Corresponds to a situation without pooling.

eTable 6: Effect estimates, Odds Ratios and type I error of studies aiming at characterizing the association between biomarker-based exposure to chemical A (ICC, 0.6) and B (ICC, 0.2) and a binary outcome, according to the number of biospecimens collected per subject and the approach used to limit the effect of exposure misclassification (1000 simulations of studies with 3000 subjects each; real effect $\beta_1 = 0$ g, assuming lack of between-assay error).

Number of biospecimens per subject	Within-subject pooling				Within-subject pooling + dissatenuation				SIMEX				Regression calibration			
	OR[a]	Effect estimate[b]	95% CI[c]	Type I error[d]	OR[a]	Effect estimate[a]	95% CI[b]	Type I error[c]	OR[a]	Effect estimate[b]	95% CI[c]	Type I error[d]	OR[a]	Effect estimate[b]	95% CI[c]	Type I error[d]
Compound A, ICC of 0.6																
1	1.00[e]	-0.002[e]	[-0.13; 0.12]	5%	1.00[e]	-0.003	[-0.22; 0.20]	5%								
2	1.01	0.004	[-0.14; 0.15]	5%	1.01	0.006	[-0.19; 0.21]	5%	1.01	0.005	[-0.18; 0.19]	6%	1.01	0.006	[-0.19; 0.21]	7%
5	1.00	-0.001	[-0.15; 0.16]	4%	1.00	-0.002	[-0.17; 0.18]	4%	1.00	-0.002	[-0.18; 0.17]	6%	1.00	-0.002	[-0.17; 0.18]	6%
10	1.00	-0.001	[-0.16; 0.16]	5%	1.00	-0.001	[-0.17; 0.17]	5%	1.00	0.000	[-0.17; 0.17]	6%	1.00	-0.001	[-0.17; 0.17]	6%
50	1.01	0.001	[-0.17; 0.16]	5%	1.01	0.001	[-0.17; 0.16]	5%	1.01	0.001	[-0.17; 0.17]	6%	1.01	0.001	[-0.17; 0.16]	7%
Compound B, ICC of 0.2																
1	1.00[e]	-0.001[e]	[-0.08; 0.07]	5%	1.01[e]	-0.007	[-0.39; 0.36]	5%								
2	1.00	0.003	[-0.09; 0.10]	6%	1.02	0.008	[-0.28; 0.30]	6%	1.01	0.004	[-0.15; 0.16]	7%	1.02	0.008	[-0.28; 0.30]	6%
5	1.00	-0.002	[-0.13; 0.12]	5%	1.00	-0.003	[-0.23; 0.21]	5%	1.00	-0.003	[-0.18; 0.17]	6%	1.00	-0.003	[-0.23; 0.21]	6%
10	1.00	-0.002	[-0.14; 0.14]	5%	1.00	-0.003	[-0.20; 0.19]	5%	1.00	-0.003	[-0.18; 0.18]	6%	1.00	-0.003	[-0.20; 0.19]	6%
50	1.00	0.001	[-0.17; 0.16]	5%	1.00	0.001	[-0.18; 0.17]	5%	1.01	0.001	[-0.18; 0.17]	6%	1.00	0.001	[-0.18; 0.17]	6%

[a] Mean of 1000 estimated odds ratio.

[b] Mean of 1000 effect estimates.

[c] Empirical confidence interval, corresponding to the empirical 2.5 and 97.5 percentiles of the health effect estimates over the 1000 simulation runs.

[d] Type I error (false positive association), proportion of studies in which the p-value of the parameter characterizing the association between the error-prone variables and health outcome was below 0.05.

[e] Corresponds to a situation without pooling

eAppendix 1: Description of the Regression Calibration and SIMEX methods[2]

Regression Calibration:

Regression calibration algorithm consists in estimating the unobserved exposure X_i (real exposure measured without error) using the repeated error-prone measurements W_{ij} and using the predicted values of X_i in the regression model estimating the association with the outcome:

$$Y_i = \beta_0 + \hat{\beta}_1 \hat{\mathrm{E}}(X_i|\overline{W_i})$$

SIMEX:

SIMEX measurement error model includes a simulation and an extrapolation step.[3] The simulation step consists of adding random pseudo error ($\sqrt{\gamma}U_{S,i,j}$) to the original values of the error-prone variable W_{ij} and fitting the regression model (continuous outcome):

$$Y_i = \alpha + \beta_{S,\gamma}\,(W_{ij} + \sqrt{\gamma}U_{S,i,j}) + \varepsilon_i$$

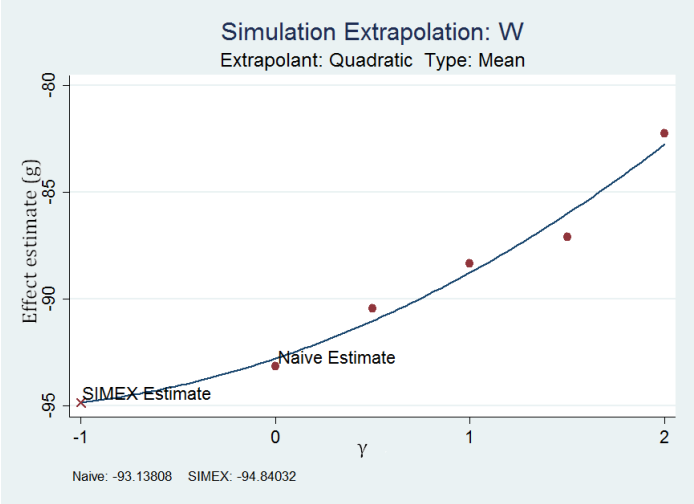
50 simulations with varying values of $U_{S,i,j}$ (normal random variable with mean, 0; SD; σ_{U_S}) are performed and the mean ($\bar{\beta}_\gamma$) of the biased estimators $\beta_{S,\gamma}$ is computed. The procedure is repeated for each value of γ in (0, 0.5, 1, 1.5 and 2).

The extrapolation step consists of modelling $\bar{\beta}_\gamma$ as a function of the measurement error variance:

$$\bar{\beta}_\gamma = \frac{{\sigma_X}^2}{{\sigma_X}^2 + {\sigma_U}^2(1+\gamma)}\widehat{\beta_1}$$

and extrapolating to the case with no measurement error (γ = -1) in order to obtain the unbiased estimator (see eFigure 3).

eFigure 3: Example of the extrapolation step of Simex using one of our simulated datasets (3000 subjects with 12 biospecimens each; real effect β_1 = -100 g, $\sigma_{Uassay}{}^2$ = 0; ICC, 0.6)



References of the Supplemental Material

1. Philippat C, Botton J, Calafat AM, Ye X, Charles MA, Slama R, Group ES. Prenatal exposure to phenols and growth in boys. *Epidemiology* 2014;**25**(5):625-35.
2. Carroll RJ, Ruppert D, Stefanski LA, eds. Measurement Error in Nonlinear Models, A Modern Perspective, Second Edition London: Chapman & Hall., 1995.
3. Cook JR, Stefanski A. Simulation-Extrapolation estimation in parametric measurement error models. . *J Am Stat Assoc* 1994;**89**:1314–1328.

```
1  ****************************************************
2  /* Exposure measurement error in biomarker based studies
3     Codes for the continous outcome, balanced desing
4    (Codes for the binary outcome are available upon request) */
5  ****************************************************
6
7  /*-------------------------------------------------------------------------------
8  Creation  : 22/11/13 (F.Perrier)
9  Modification : 21/10/14 (C.Philippat)
10 Modification : 25/03/15 (F.Perrier)
11 Modification : 28/05/2015 (F. Perrier)
12 Software used: STATA/SE, version 13 (StataCorp, College Station, TX, USA)
13
14 Package that need to be install before running the program :
15 net from http://www.stata.com/merror
16 net install merror // package needed in order to use simex and rcal fonctions
17
18 -------------------------------------------------------------------------------*/
19 * version 12.1
20
21 cap program drop _all
22 clear all
23 *set matsize 6000
24 set more off
25
26 timer on 1
27
28 /* Definitions of the paths to folders in which we will store the
29 simulated datasets and the results of the simulation */
30 ********************************************************************************
31 cd "D:\Erreur de mesure"
32 global simulation = "Simulations\reg_lin_test_25-03-15" /*dataset storage */
33 global results = "Resultats\reg_lin_test_25-03-15"     /*result storage */
34
35 /* Definition of the parameters used in the simulation */
36 ********************************************************************************
37 global b0 =  14900           /* Intercept of the regression model */
38 global b1 = -100 /*0*/       /* Effect of the real exposure (X) on the outcome (Y) */
39 global nb_indiv = 3000       /*500 1000 2000 4000*/  /* Number of subjects in the simulated population */
40 global nb_simul = 2       /* Number of repetitions / simulations */
41 global corr_list = "0.6 0.2" /* Intraclass correlation coefficient observed between repeated measurements */
42 global sample_list = "1 2 3 4 5 6 7 8 9 10 12 15 18 20 25 30 35 40 50"  /* List containing  the different
43 numbers of biospecimens selected for each participants to study the association with the health outocme */
44 global var_Uassay_list "0 0.01 0.1 0.5" /* Variance of between-assay error */
45
46 /*-------------------------------------------------------------------------------------------------------------------------------*/
47 * Folder creation (needed datasets and results storage)
48 * mkdir "...."
49 /*
50 * datasets
51 *************
52 foreach c in $corr_list{
53    mkdir "$simulation\Correlation `c'"
54 }
55
56 foreach b in $b1 0 {
57    foreach c in $corr_list {
58       mkdir "$simulation\Correlation `c'\beta `b'"
59    }
60 }
61
62 foreach n in  500 1000 2000 3000 4000 {
63    foreach b in $b1 0 {
64       foreach c in $corr_list {
65          mkdir "$simulation\Correlation `c'\beta `b'\n `n'"
66       }
67    }
68 }
69
70 * Results
71 ************
72
73 foreach c in $corr_list {
74    mkdir "$results\Correlation `c'"
75 }
76
77 foreach b in $b1 0 {
78    foreach c in $corr_list {
79       mkdir "$results\Correlation `c'\beta `b'"
80    }
81 }
82
83 foreach n in  500 1000 2000 3000 4000 {
84    foreach b in $b1 0 {
85       foreach c in $corr_list {
86          mkdir "$results\Correlation `c'\beta `b'\n `n'"
87       }
```

```
   }
}

*/

/*-------------------------------------------------------------------*/

/* ************************************************************
DATA SET SIMULATION
In this section we simulated the datasets
* Notations:
X: real exposure
Uwithin: within-subject error
Uassay: between-assay error
W: Exposures measured with error
Y: Continuous health outcome
**********************************************************************/

foreach var_Ua in $var_Uassay_list{
   global var_Uassay=`var_Ua'   /* Variance of the between-assay error */
   di in green "variance between-assay : $var_Uassay"

   foreach corr in $corr_list{
      set seed 18102013
      global c = `corr'   /* ICC */
      di in green "  Data set simulation for beta:$b1 ICC:$c"

      forvalues i = 1/$nb_simul {
         di in yellow "Simulation number : " `i'  " / " $nb_simul

         foreach j in $sample_list{
            /* Simulation of X (True exposure) */
               global fixed_mean_X = 0   // mean
               global fixed_var_X =  1   // variance
               mat m_X = $fixed_mean_X
               mat sd_X = sqrt($fixed_var_X )
               qui drawnorm X, mean(m_X) sds(sd_X) n($nb_indiv)
               qui sum X
               global var_X_true = r(Var)


            /* Simulation of Uassay (inter-assay error) */
               mat m_Uassay = J(1,`j',0)
               mat sd_Uassay = J(1,`j',sqrt($var_Uassay))
               if `j'==1{
                  qui drawnorm Uassay_1,  mean(m_Uassay) sds(sd_Uassay) n($nb_indiv)
               }
               else {
                  qui drawnorm Uassay_1-Uassay_`j',  mean(m_Uassay) sds(sd_Uassay) n($nb_indiv)
               }

            /* Simulation of Uwithin (within-subject error) */
               global var_Uwithin = $var_X_true * (1/$c -1)   // variance of error Uwitihn
               mat m_Uwithin = J(1,`j',0)
               mat sd_Uwithin = J(1,`j',sqrt($var_Uwithin))
               if `j'==1{
                  qui drawnorm Uwithin_1,  mean(m_Uwithin) sds(sd_Uwithin) n($nb_indiv)
               }
               else {
                  qui drawnorm Uwithin_1-Uwithin_`j',  mean(m_Uwithin) sds(sd_Uwithin) n($nb_indiv)
               }

            /* Computation of the concentration measured in the pool */
               qui egen mean_Uwithin = rowmean(Uwithin_1-Uwithin_`j')  /* Mean of the within-subject error
               error */
               qui gen pool = X + mean_Uwithin + Uassay_1  /* Concentration measured in the pool */

            /* Simulation of the concentration measured in each biospeciment (W) */
               forvalues k = 1/`j' {
                  qui gen W_`k' = X + Uwithin_`k' + Uassay_`k'
               }


            /* Simulation of the continuous health outcome */
               qui drawnorm E,  mean(0) sds(1650) n($nb_indiv) /* Simulation of the model error,
               sds(1650) is the SD observed in the French mother child cohort EDEN, for the
               weight at 3 years */
               qui gen Y = $b0 + $b1*X + E
               qui save "$simulation/Correlation 0$c/beta $b1/n
$nb_indiv/simul`i'_`j'samples_n${nb_indiv}_varUass${var_Uassay}.dta", replace

               drop *
         }
      }
   }


```

```
/*-------------------------------------------------------------------------------------------------------------*/

/* ********************************************************************************************
 ANALYSES
 In this section we studied the associations between the outcome (Y) and the real exposure (1) or the
   error prone variable (2) using 3 different approaches: within-subject pooling (2a), Regression
   Calibration (2b), Simex (2c). The a posteriori disattenuation is presented line 283.
**********************************************************************************************/

   foreach corr in $corr_list{
      global c = `corr'    /* ICC */
      di in green "  Analyses for beta:$b1 ICC:$c"

      forvalues i = 1/$nb_simul {
         di in yellow "Simulation number  : " `i'  " / " $nb_simul

         local it = 0 /* sample_list iterator */
         foreach j in $sample_list{
            local it = `it' + 1
            use  "$simulation/Correlation 0$c/beta $b1/n
$nb_indiv/simul`i'_`j'samples_n${nb_indiv}_varUass${var_Uassay}.dta", clear

      /* 1- Calculation and storage of the real effect;  beta coefficients obtained when we used the
            real exposure X */
            qui regress Y X
            mat table_true = r(table)
            global beta = table_true[1,1]       // regression coefficient
            global p = table_true[4,1]          // p-value
            mata : res_temp_true = (`j', $beta, $p)

      /* 2a- Pooling method */
            qui regress Y pool
            mat table_pool = r(table)
            global beta = table_pool[1,1]       // regression coefficient
            global se = table_pool[2,1]         // standard error
            global sd = $se * sqrt($nb_indiv)   // computation of the standard deviation
            global p = table_pool[4,1]          // p-value
            mata :   res_temp_pool = (`j', $beta, $sd, $p)

            if (`j'==1){
               mata : res_temp_simex = (`j', ., .)
               mata : res_temp_rcal  = (`j', ., .)
            }
            else{

         /* 2b- SIMEX */
               qui simex(Y=)(W : W_1-W_`j'), bstrap brep(100) seed(1)
               mat tab_simex = e(theta)
               global  beta = tab_simex[1,2]       //regression coefficient
               mat tab_simex = e(V)
               global se = sqrt(tab_simex[1,1])    // standard error
               global sd = $se * sqrt($nb_indiv)   // computation of the standard deviation
               global t = $beta/$se                // student test (hypothesis : beta=0)
               global p = 2*ttail(e(df_r),abs($t)) // computation of the p-value
               /* note: For large sample size the student distribution could be approximated by a normal distribution,
               however for small sample size the normal distribution should be used to compute p-values */
               // e(df_r) : degree of freedom
               // inverse Student distribution
               mata : res_temp_simex = (`j', $beta, $p)

         /* 2c- Regression Calibration */
               qui rcal (Y=) (W : W_1-W_`j'), bstrap brep(100) seed(1)
               mat tab_rcal = e(b)
               global  beta = tab_rcal[1,1]        // regression coefficient
               mat tab_rcal = e(V)
               global se = sqrt(tab_rcal[1,1])     // standard error
               global sd = $se * sqrt($nb_indiv)   // computation of the standard deviation
               global t = $beta/$se                // student test (hypothesis : beta=0)
               global p = 2*ttail(e(df_r),abs($t)) // computation of the p-value
               /* note: For large sample size the student distribution could be approximated by a normal distribution,
               however for small sample size the normal distribution should be used to compute p-values */

               mata : res_temp_rcal = (`j', $beta, $p)
            }

         /* Storage of the results obtained with the different methods in mata matrix */
            if `i'==1 & `it'==1 {
                  mata : res_all_true = res_temp_true
                  mata : res_all_pool = res_temp_pool
                  mata : res_all_simex = res_temp_simex
                  mata : res_all_rcal = res_temp_rcal
            }
            else {
                  mata : res_all_true = res_all_true\res_temp_true
                  mata : res_all_pool = res_all_pool\res_temp_pool
                  mata : res_all_simex = res_all_simex\res_temp_simex
                  mata : res_all_rcal = res_all_rcal\res_temp_rcal
```

```
260             }
261          }
262       }
263
264       clear
265
266   /* ------------------------------------------------------------------------------------------*/
267
268   /*****************************************************************************
269   COMPUTATION OF THE RESULTS OBTAINED AFTER RUNNING THE 1000 SIMULATIONS AND STORAGE
270   *****************************************************************************/
271
272      /* Transformation of the matrix into variables */
273
274         // Results obtained with the real exposure
275         getmata (nb_ech all_beta_true all_p_val_true) = res_all_true, replace force
276         // Results obtained with the within-subject pooling method
277         getmata (nb_ech all_beta_pool all_sd_pool all_p_val_pool) = res_all_pool, replace force
278         // SIMEX
279         getmata (nb_ech all_beta_simex all_p_val_simex) = res_all_simex, replace force
280         // Regression Calibration
281         getmata (nb_ech all_beta_rcal all_p_val_rcal) = res_all_rcal, replace force
282
283
284      /* 2d- A posteriori disattenuation of the effect estimates obtained with pooling  */
285         gen all_beta_corr = all_beta_pool * ( (nb_ech-1)/nb_ech + 1/(nb_ech * $c) )
286         gen all_sd_corr = all_sd_pool * ( (nb_ech-1)/nb_ech + 1/(nb_ech * $c) )
287         gen all_IC_inf_corr = all_beta_corr - 1.96 * all_sd_corr / sqrt($nb_indiv)
288         gen all_IC_sup_corr = all_beta_corr + 1.96 * all_sd_corr / sqrt($nb_indiv)
289
290      /* Results computation for each number of samples `j' used per participant  */
291         local it = 0 /* sample_list iterator */
292
293         foreach j in $sample_list{
294            local it = `it' + 1
295            /* Mean and variance */
296            qui sum all_beta_true if nb_ech==`j'
297            global mean_beta_true = r(mean)        // betas true exposure
298            global sd_beta_true = r(sd)
299
300            qui sum all_beta_pool if nb_ech==`j'
301            global mean_beta_pool = r(mean)        // betas pooling
302            global sd_beta_pool = r(sd)
303
304            qui sum all_beta_corr if nb_ech==`j'
305            global mean_beta_corr = r(mean)   /* betas pooling +
306            a posteriori disattenuation */
307            global sd_beta_corr = r(sd)
308
309            if (`j'==1) {
310               global mean_beta_simex = .
311               global mean_beta_rcal = .
312
313               global sd_beta_simex = .
314               global sd_beta_rcal = .
315
316               global puiss_simex   = .
317               global puiss_rcal = .
318
319            }
320            else {
321               qui sum all_beta_simex if nb_ech==`j'
322               global mean_beta_simex = r(mean)    // betas SIMEX
323               global sd_beta_simex = r(sd)
324
325               qui sum all_beta_rcal if nb_ech==`j'
326               global mean_beta_rcal = r(mean)     // betas regression calibration
327               global sd_beta_rcal = r(sd)
328            }
329
330            /* Empirical Confidence Intervals */
331            qui centile (all_beta_pool) if nb_ech==`j', centile (2.5 97.5)  // pooling
332            global IC_inf_pool = r(c_1)
333            global IC_sup_pool = r(c_2)
334
335            qui centile (all_beta_corr) if nb_ech==`j', centile (2.5 97.5)  /* pooling + a
336            posteriori disattenuation */
337            global IC_inf_corr = r(c_1)
338            global IC_sup_corr = r(c_2)
339
340            qui centile (all_beta_simex) if nb_ech==`j', centile (2.5 97.5) // SIMEX
341            global IC_inf_simex = r(c_1)
342            global IC_sup_simex = r(c_2)
343
344            qui centile (all_beta_rcal) if nb_ech==`j', centile (2.5 97.5) // regression calibration
345            global IC_inf_rcal = r(c_1)
346            global IC_sup_rcal = r(c_2)
```

```
347
348           /* Power (percentage of p-value < 0.05)*/
349           qui count if all_p_val_true<0.05 & nb_ech==`j'      // true exposure
350           global puiss_true = r(N)/$nb_simul
351
352           qui count if all_p_val_pool<0.05 & nb_ech==`j'      // pooling
353           global puiss_pool = r(N)/$nb_simul
354
355           qui count if all_IC_inf_corr<0 & all_IC_sup_corr>0 & nb_ech==`j' /* pooling +
356           a posteriori disattenuation */
357           global puiss_corr = (${nb_simul}-r(N)) / $nb_simul
358
359           if (`j'>1) {
360              qui count if all_p_val_simex<0.05 & nb_ech==`j' // SIMEX
361              global puiss_simex   = r(N)/$nb_simul
362              qui count if all_p_val_rcal<0.05 & nb_ech==`j'  // regression calibration
363              global puiss_rcal = r(N)/$nb_simul
364           }
365
366     /* CREATION OF MATRIX RESULT*/
367           mat indiv_beta = `j', $mean_beta_true, $mean_beta_pool,  $mean_beta_corr,  $mean_beta_simex,  $mean_beta_rcal
368           mat indiv_sd = `j',  $sd_beta_true, $sd_beta_pool, $sd_beta_corr, $sd_beta_simex, $sd_beta_rcal
369           mat indiv_IC_inf = `j', $IC_inf_pool, $IC_inf_corr, $IC_inf_simex,  $IC_inf_rcal
370           mat indiv_IC_sup = `j', $IC_sup_pool, $IC_sup_corr, $IC_sup_simex,  $IC_sup_rcal
371           mat indiv_puiss = `j', $puiss_true, $puiss_pool, $puiss_corr, $puiss_simex, $puiss_rcal
372
373           if `it'==1 {   // creation of the first line of result tables
374              mat res_beta = indiv_beta       // Mean of the estimators
375              mat res_sd = indiv_sd           // Standard deviation of the estimators
376              mat res_IC_inf = indiv_IC_inf   // Lower bound of the CI of the estimators
377              mat res_IC_sup = indiv_IC_sup   // Upper bound of the CI of the estimators
378              mat res_puiss = indiv_puiss     // Power of the estimators
379           }
380           else {      // add of a line on result tables already existing
381              mat res_beta = res_beta\indiv_beta
382              mat res_sd = res_sd\indiv_sd
383              mat res_IC_inf = res_IC_inf\indiv_IC_inf
384              mat res_IC_sup = res_IC_sup\indiv_IC_sup
385              mat res_puiss = res_puiss\indiv_puiss
386           }
387        }
388
389        /* Rename the name of the matrix column*/
390        matrix colnames res_beta = "p" "True" "Pooling" "Pooling + Disatt" "SIMEX" "Reg-Cal"
391        matrix colnames res_sd = "p" "True" "Pooling" "Pooling + Disatt" "SIMEX" "Reg-Cal"
392        matrix colnames res_IC_inf = "p" "Pooling" "Pooling + Disatt" "SIMEX" "Reg-Cal"
393        matrix colnames res_IC_sup = "p" "Pooling" "Pooling + Disatt" "SIMEX" "Reg-Cal"
394        matrix colnames res_puiss = "p" "True" "Pooling" "Pooling + Disatt" "SIMEX" "Reg-Cal"
395
396     /* ------------------------------------------------------------------------------*/
397     /* MATRIX TRANSFORMATION IN VARIABLES + RENAME OF THESE VARIABLES*/
398
399        /* Mean of betas*/
400        svmat res_beta
401        rename res_beta1 p
402        rename res_beta2 beta_true
403        rename res_beta3 beta_pool
404        rename res_beta4 beta_corr
405        rename res_beta5 beta_simex
406        rename res_beta6 beta_rcal
407
408        /* Percentages of bias in the effect estimates compared to the real effect */
409        qui gen perc_bias_pool = abs((beta_pool - beta_true) / beta_true) * 100
410        qui gen perc_bias_corr = abs((beta_corr - beta_true) / beta_true)* 100
411        qui gen perc_bias_simex = abs((beta_simex - beta_true) / beta_true)* 100
412        qui gen perc_bias_rcal = abs((beta_rcal - beta_true) / beta_true) * 100
413
414        /* Variance of betas*/
415        svmat res_sd
416        drop res_sd1
417        rename res_sd2 sd_true
418        rename res_sd3 sd_pool
419        rename res_sd4 sd_corr
420        rename res_sd5 sd_simex
421        rename res_sd6 sd_rcal
422
423        /* Confidence interval*/
424           /* Lower bound*/
425        svmat res_IC_inf
426        drop res_IC_inf1
427        rename res_IC_inf2 IC_inf_pool
428        rename res_IC_inf3 IC_inf_corr
429        rename res_IC_inf4 IC_inf_simex
430        rename res_IC_inf5 IC_inf_rcal
431
432           /* Upper bound*/
433        svmat res_IC_sup
```

```
434         drop res_IC_sup1
435         rename res_IC_sup2 IC_sup_pool
436         rename res_IC_sup3 IC_sup_corr
437         rename res_IC_sup4 IC_sup_simex
438         rename res_IC_sup5 IC_sup_rcal
439
440         /* Power*/
441         svmat res_puiss
442         drop res_puiss1
443         rename res_puiss2 puiss_true
444         rename res_puiss3 puiss_pool
445         rename res_puiss4 puiss_corr
446         rename res_puiss5 puiss_simex
447         rename res_puiss6 puiss_rcal
448
449         /* Round of variables */
450         global var_list = "pool corr simex rcal"
451
452         qui replace beta_true = round(beta_true,1)
453         qui replace puiss_true = round(puiss_true,0.01)
454
455         foreach var in $var_list {
456            qui replace beta_`var' = round(beta_`var',1)
457            qui replace sd_`var' = round(sd_`var',0.1)
458            qui replace IC_inf_`var' = round(IC_inf_`var',0.1)
459            qui replace IC_sup_`var' = round(IC_sup_`var',0.1)
460            qui replace puiss_`var' = round(puiss_`var',0.01)
461            qui replace perc_bias_`var' = round(perc_bias_`var',1)
462
463            qui tostring(IC_inf_`var'), gen(string_IC_inf_`var') force format(%7.0f)
464            qui tostring(IC_sup_`var'), gen(string_IC_sup_`var') force format(%7.0f)
465            qui gen IC_`var' = "[" + string_IC_inf_`var' + " ; " + string_IC_sup_`var' + "]" if string_IC_inf_`var' !=
    "." & string_IC_sup_`var' != "."
466         }
467
468         /* Back up of data set containing results */
469         save "$results/Correlation 0$c/beta $b1/n $nb_indiv/res_n${nb_indiv}_varUass${var_Uassay}_sim${nb_simul}.dta",
    replace
470
471      preserve
472          /* Exportation of results on Excel */
473          keep p beta_true puiss_true beta_pool IC_pool puiss_pool perc_bias_pool ///
474            beta_corr IC_corr puiss_corr perc_bias_corr ///
475            beta_simex IC_simex puiss_simex perc_bias_simex ///
476            beta_rcal IC_rcal puiss_rcal perc_bias_rcal
477         export excel using "$results/Correlation 0$c/beta $b1/n
    $nb_indiv/tab_corr0${c}_beta${b1}_n${nb_indiv}_varUass${var_Uassay}_sim${nb_simul}.xls", firstrow(variables) replace
478      restore
479
480
    /*-------------------------------------------------------------------------------------------------------------------
    -----------*/
481
482         /* GRAPHICS : ARTICLE - FIGURE 1 */
483
484         /* Percentage bias */
485         twoway (line perc_bias_pool p, lcolor(black) lpattern(solid)) ///
486           (line perc_bias_corr p, lcolor(black) lpattern(dot)) ///
487           (line perc_bias_simex p, lcolor(black) lpattern(shortdash))  ///
488           (line perc_bias_rcal p, lcolor(black) lpattern(dash))  if p>=2 & p<=30, ///
489           xtitle("Number of urinary samples per subject", size(medlarge)) xlabel(2 5 10 15 20 25 30) ///
490           ytitle("Bias in the effect estimate (%)", size(medlarge))  legend(label(1 "Pooling method") ///
491           label(2 "Pooling + Disattenuation")  label(3 "Simex") label(4 "Regression-Calibration") cols(2) span bmargin(
    zero) ) ///
492           name(graph_bias, replace) scheme(s2gmanual) graphregion(color(white)) bgcolor(white)
493
494         /* BACK UP OF GRAPHICS */
495         qui graph save graph_bias "$results/Correlation 0$c/beta $b1/n
    $nb_indiv/graph_bias_n${nb_indiv}_varUass${var_Uassay}_sim${nb_simul}.gph", replace
496
497      }
498      drop *
499
500   }
501
502   timer off 1
503   timer list 1
504
505
506
507
508
```

```
1  ****************************************************
2  /* Exposure measurement error in biomarker based studies
3     Codes for the continous outcome, unbalanced desing
4    (Codes for the binary outcome are available upon request) */
5  ****************************************************
6
7     /*-------------------------------------------------------------------------------------------------------
8     Creation  : 15/04/14 (F.Perrier)
9     Modification : 28/05/15 (F.Perrier)
10    Software used: STATA/SE, version 13 (StataCorp, College Station, TX, USA)
11
12    Package that need to be install before running the program :
13    net from http://www.stata.com/merror
14    net install merror // package needed to use simex and regression calibration fonctions
15
16    *Note: This program does not take into account between-assay error
17    -------------------------------------------------------------------------------------------------------*/
18
19    cap program drop _all
20    clear all
21    *set matsize 6000
22    set more off
23
24    timer clear
25    timer on 1
26
27    /* Definitions of the paths to folders in which we will store the simulated datasets and the
28    results of the simulation */
29    ********************************************************************************************************
30    cd "D:\Erreur de mesure"
31    global simulation = "Simulations\nb_ech_diff_lin_22-10-14" /*dataset storage */
32    global results = "Resultats\nb_ech_diff_lin_22-10-14"  /*result storage */
33
34    /* Definition of the parameters used in the simulation */
35    ********************************************************************************************************
36    global b0 = 14900        /* Intercept of the linear regression model */
37    global b1 = -100         /* Effect of the real exposure (X) on the outcome (Y) */
38
39    global nb_indiv = 3000   /* Number of subjects in the simulated population */
40    global nb_simul = 1000   /* Number of repetitions / simulations */
41    global nb_ech_tot = 9000 /* Number total of biospecimens */
42    global corr_list = "0.6 0.2" /* Intraclass correlation coefficient observed between repeated
43    measurements of exposure */
44
45    /*Definition of the different scenarios to simulate */
46    ********************************************************************************************************
47    /* the following study desing will be simulated
48       Scenario a : 600 subjects with 1 sample  / 1500 with 2 samples / 900 with 6 samples
49       Scenario b : 600 subjects with 1 sample  / 1800 avec 3 samples / 600 with 5 samples
50       Scenario c : 900 subjects with 1 sample  / 1200 with 3 samples / 900 with 5 samples
51       Scenario d : 1200 subjects with 1 sample / 1200 with 2 samples / 600 with 9 samples
52    */
53    global nb_sce = 4          /* Number of different designs to test */
54    mat z_1 = 1\1\1\1          /* Number of biospecimens to simulate for the group 1 for each design */
55    mat z_2 = 2\3\3\2          /* Number of biospecimens to simulate for the group 2 for each design  */
56    mat z_3 = 6\5\5\9          /* Number of biospecimens to simulate for the group 3 for each design  */
57    mat p_1 = 0.2\0.2\0.3\0.4  /* Percentage of the population belonging to group 1 */
58    mat p_2 = 0.7\0.8\0.7\0.8  /* Cumulative percentage of the population belonging to group 1 or 2*/
59
60    /*-------------------------------------------------------------------------------------------------------*/
61    * Folder creation (needed datasets and results storage)
62    * mkdir "...."
63    /*
64    * datasets
65    *************
66    forvalue k = 1/$nb_sce{
67       mkdir "$simulation\Scenario_`k'"
68    }
69
70    foreach c in $corr_list {
71       forvalue k = 1/$nb_sce{
72          mkdir "$simulation\Scenario_`k'\ICC_`c'"
73       }
74    }
75    */
76    /*-------------------------------------------------------------------------------------------------------*/
77
78    /* ********************************************************************************************************
79    DATA SET SIMULATION /* We assumed a lack of between-assay error */
80    In this section we simulated the datasets
81    Notations:
82    X: Real exposure
83    Uwithin: within-subject error
84    W: Exposure measure with error
85    Y: Continuous health outcome
86    ******************************************************************************************** */
87    set obs $nb_indiv
```

```
88      set seed 15042014
89
90      /*-------------------------------------------------------------------------------------------------------*/
91      local it = 0  /* simulations iterator  */
92
93      forvalue k = 1/$nb_sce {
94
95         global z_1 = z_1[`k',1]   /* Number of samples to simulate for the group 1 for the kth scenario */
96         global z_2 = z_2[`k',1]   /* Number of samples to simulate for the group 2 for the kth scenario */
97         global z_3 = z_3[`k',1]   /* Number of samples to simulate for the group 3 for the kth scenario */
98         global p_1 = p_1[`k',1]   /* Percentage of the population belonging to group 1 for the kth
99      scenario */
100        global p_2 = p_2[`k',1]   /* Cumulative percentage of the population belonging to groups 1 or
101     2 for the kth scenario */
102
103        di in red " Scenario " `k' " :  - " $p_1*100  "% of the population with " $z_1 " sample(s) "
104        if ($p_1!=1 & $z_2!=0) {
105           di in red "      - " ($p_2-$p_1)*100 "% of the population with " $z_2 " samples "
106        }
107        if ($p_2!=1 & $z_3!=0) {
108           di in red "      - " (1-$p_2)*100 "% of the population with " $z_3 " samples "
109        }
110
111        foreach corr in $corr_list {
112           global c = `corr'
113           di in green "  ICC 0" $c
114
115           forvalue i = 1/$nb_simul {
116              local it = `it' + 1
117              if (mod(`i',$nb_simul/10)==0) {
118                 di in yellow   `i'/($nb_simul/10) "/10 of data are simulated "
119              }
120
121              qui set obs $nb_indiv
122              gen id = _n
123
124              /* Simulation of X (True exposure) */
125              global fixed_mean_X = 0            // mean
126              global fixed_var_X =  1               // variance
127              mat m_X = $fixed_mean_X
128              mat sd_X = sqrt($fixed_var_X )
129              qui drawnorm X, mean(m_X) sds(sd_X) n($nb_indiv)
130
131              /* Simulation of Uwithin (within-subject error)
132              / We assumed a lack of between-assay error */
133              qui sum X
134              global var_X_true = r(Var)
135              global var_Uwithin_true = $var_X_true * (1/$c -1)  // fixed variance of error
136              mat m_Uwithin = J(1,$z_1,0)
137              mat sd_Uwithin = J(1,$z_1,sqrt($var_Uwithin_true))
138              global temp = $z_1 + 1
139
140              if $z_1 == 1 { /* Group 1 */
141                 qui drawnorm Uwithin_1,  mean(m_Uwithin) sds(sd_Uwithin) n($nb_indiv)
142              }
143              else{
144                 qui drawnorm Uwithin_1-Uwithin_$z_1 ,  mean(m_Uwithin) sds(sd_Uwithin) n($nb_indiv)
145              }
146
147
148              if $z_2!=0 { /* Group 2 */
149                 forvalues j = $temp/$z_2 {
150                    qui gen Uwithin_`j' = rnormal(0,sqrt($var_Uwithin_true)) if id>$p_1*$nb_indiv
151                 }
152              }
153
154              global temp = $z_2 + 1
155              if $z_3!=0 { /* Group 3 */
156                 forvalues j = $temp/$z_3 {
157                    qui gen Uwithin_`j' = rnormal(0,sqrt($var_Uwithin_true)) if id>($p_2)*$nb_indiv
158                 }
159              }
160
161              /* Simulation of W (exposures measured with error) */
162              global z_max = max($z_1, $z_2, $z_3)
163              forvalues j = 1/$z_max {
164                 qui gen W_`j' = X + Uwithin_`j'
165              }
166
167              /* Simulation of the continuous health outcome */
168              qui drawnorm E,  mean(0) sds(1650) n($nb_indiv)  /* Simulation of the model error,
169                 sds(1650) is the SD observed in the French mother child cohort EDEN, for the
170                 weight at 3 years */
171
172              qui gen Y = $b0 + $b1*X + E
173
174              if (max($p_1, $p_2) == $p_1 & $p_1 != 1) {
```

```
                qui sample 50
                qui replace id = _n
             }

             qui save "${simulation}\Scenario_`k'\ICC_0${c}\simul`i'_beta${b1}_n${nb_indiv}", replace


    /* ************************************************************************************************
    ANALYSES
    In this section we study the associations between the outcome (Y) and the real exposure (1) or the
    error prone variable using different approaches: Pooling (2), Regression Calibration (3) and  Simex (4)
    ************************************************************************************************ */


             /* 1- Calculation and storage of beta true, the beta coefficient we obtained when we used
             the real exposure X */
             qui regress Y X
             mat table_true = r(table)
             global beta = table_true[1,1]     //regression coefficient
             global se = table_true[2,1]       //standard error
             global sd = $se * sqrt($nb_indiv) //standard deviation
             global puiss = table_true[4,1]    //p-value
             mata : res_temp_true = (`k', $c, $beta, $sd, $puiss)

             /* 2a- Pooling */
             qui egen W_pool = rowmean(W_1-W_$z_max)   /* computation of the mean of the repeated measures
             for each individual  */
             qui gen weight = $z_1  // definition of a weight for each subject
             if (max($z_1, $z_2)!= $z_1) {
                qui replace weight = $z_2 if id > ($p_1)*$nb_indiv & id <= ($p_2)*$nb_indiv
                qui replace weight = $z_3 if id > ($p_2)*$nb_indiv
             }
             qui regress Y W_pool [aw=weight]

             mat table_pool = r(table)
             global beta = table_pool[1,1]     //regression coefficient
             global se = table_pool[2,1]       //standard error
             global sd = $se * sqrt($nb_indiv) //standard deviation
             global puiss = table_pool[4,1]    //p-value
             mata :   res_temp_pool = (`k', $c, $beta, $sd, $puiss)


             /* 3- SIMEX */
             qui simex(Y=)(W : W_1-W_$z_max), bstrap brep(100) seed(1)

             mat tab_simex = e(theta)
             global  beta = tab_simex[1,2]       //regression coefficient
             mat tab_simex = e(V)
             global se = sqrt(tab_simex[1,1])    //standard error
             global sd = $se * sqrt($nb_indiv)   //standard deviation
             global t = $beta/$se                //student test (hypothesis : beta=0)
             global puiss = 2*ttail(e(df_r),abs($t))       //computation of the p-value
             // e(df_r) : degree of freedom
             // inverse distribution of Student
             /* note: For large sample size the student distribution could be approximated by a normal
                distribution, however for small sample size the normal distribution should be used to
                compute p-values */
             mata : res_temp_simex = (`k', $c, $beta, $sd, $puiss)

             /* 4- Regression-Calibration */
             qui rcal (Y=) (W : W_1-W_$z_max), bstrap brep(100) seed(1)

             mat tab_rcal = e(b)
             global  beta = tab_rcal[1,1]        //regression coefficient
             mat tab_rcal = e(V)
             global se = sqrt(tab_rcal[1,1])     //standard error
             global sd = $se * sqrt($nb_indiv)   //standard error
             global t = $beta/$se                //student test (hypothesis : beta=0)
             global puiss = 2*ttail(e(df_r),abs($t))       //computation of the p-value
             /* note: For large sample size the student distribution could be approximated by a normal
                distribution, however for small sample size the normal distribution should be used to
                compute p-values */
             mata : res_temp_rcal = (`k', $c, $beta, $sd, $puiss )

             /* Storage of the results obtained with the different methods in mata matrix */
             if `it'==1 {
                mata : res_all_true = res_temp_true
                mata : res_all_pool = res_temp_pool
                mata : res_all_simex = res_temp_simex
                mata : res_all_rcal = res_temp_rcal
             }
             else {
                mata : res_all_true = res_all_true\res_temp_true
                mata : res_all_pool = res_all_pool\res_temp_pool
                mata : res_all_simex = res_all_simex\res_temp_simex
                mata : res_all_rcal = res_all_rcal\res_temp_rcal
             }
```

```
            drop *
         }
      }
   }

   /*-------------------------------------------------------------------------------------------------------*/

   /* ******************************************************************************************************
   COMPUTATION OF THE RESULTS OBTAINED AFTER RUNNING THE 1000 SIMULATIONS AND STORAGE
   ****************************************************************************************************** */

   /* Transformation of the matrix into variables */

   //Beta true
   getmata (all_num_sce all_ICC all_beta_true all_sd_true all_p_val_true) = res_all_true, replace force
   //Pooling method
    getmata (all_num_sce all_ICC all_beta_pool all_sd_pool all_p_val_pool) = res_all_pool, replace force
   //SIMEX
   getmata (all_num_sce all_ICC all_beta_simex all_sd_simex all_p_val_simex) = res_all_simex, replace force
   //Regression Calibration
   getmata (all_num_sce all_ICC all_beta_rcal all_sd_rcal all_p_val_rcal) = res_all_rcal, replace force

   /*-------------------------------------------------------------------------------------------------------*/
   /* Results computation for each number of samples `j' used per participant (n = 1000 simulations) */
   local it = 0

   forvalue k = 1/$nb_sce {
      global z_1 = z_1[`k',1]   /* Number of samples simulated for the group 1 for the kth scenario */
      global z_2 = z_2[`k',1]   /* Number of samples simulated for the group 2 for the kth scenario */
      global z_3 = z_3[`k',1]   /* Number of samples simulated for the group 3 for the kth scenario */
      global p_1 = p_1[`k',1]   /* Percentage of the population belonging to group 1 for the kth
   scenario */
      global p_2 = p_2[`k',1]   /* Cumulative percentage of the population belonging to groups 1 or 2
   for the kth scenario */

      foreach corr in $corr_list {
         global c = `corr'
         local it = `it' + 1

         /* Mean and variance */
         qui sum all_beta_true if all_num_sce==`k' & all_ICC==$c
         global mean_beta_true = r(mean)        //betas true
         global sd_beta_true = r(sd)

         qui sum all_beta_pool if all_num_sce==`k' & all_ICC==$c
         global mean_beta_pool = r(mean)        //pooling method
         global sd_beta_pool = r(sd)

         qui sum all_beta_simex if all_num_sce==`k' & all_ICC==$c
         global mean_beta_simex = r(mean)       //Simex
         global sd_beta_simex = r(sd)

         qui sum all_beta_rcal if all_num_sce==`k' & all_ICC==$c
         global mean_beta_rcal = r(mean)        //Regression calibration
         global sd_beta_rcal = r(sd)

         /* Confidence Interval of estimators*/
         qui centile (all_beta_pool) if all_num_sce==`k' & all_ICC==$c, centile (2.5 97.5)
         global IC_inf_pool = r(c_1)   //pooling method
         global IC_sup_pool = r(c_2)

         qui centile (all_beta_simex) if all_num_sce==`k' & all_ICC==$c, centile (2.5 97.5)
         global IC_inf_simex = r(c_1)  //Simex
         global IC_sup_simex = r(c_2)

         qui centile (all_beta_rcal) if all_num_sce==`k' & all_ICC==$c, centile (2.5 97.5)
         global IC_inf_rcal = r(c_1)   //Regression calibration

         global IC_sup_rcal = r(c_2)

         /* Power (percentage of p-value upper than 0.05 => acceptance percentage of H0 (beta=0) )*/
         qui count if all_p_val_true<0.05 & all_num_sce==`k' & all_ICC==$c     //true exposure
         global puiss_true = r(N)/$nb_simul

         qui count if all_p_val_pool<0.05 & all_num_sce==`k' & all_ICC==$c     //pooling method
         global puiss_pool = r(N)/$nb_simul

         qui count if all_p_val_simex<0.05 & all_num_sce==`k' & all_ICC==$c    //Simex
         global puiss_simex   = r(N)/$nb_simul

         qui count if all_p_val_rcal<0.05 & all_num_sce==`k' & all_ICC==$c     //Regression calibration
         global puiss_rcal = r(N)/$nb_simul

      /* CREATION OF MATRIX RESULT*/
         mat indiv_beta = `k', $c, $mean_beta_true, $mean_beta_pool, $mean_beta_simex, $mean_beta_rcal
         mat indiv_sd = `k', $c,  $sd_beta_true, $sd_beta_pool, $sd_beta_simex, $sd_beta_rcal
         mat indiv_IC_inf = `k', $c, $IC_inf_pool, $IC_inf_simex, $IC_inf_rcal
```

```
            mat indiv_IC_sup = `k', $c, $IC_sup_pool, $IC_sup_simex, $IC_sup_rcal
            mat indiv_puiss = `k', $c, $puiss_true, $puiss_pool, $puiss_simex, $puiss_rcal

            if `it'==1 {   //creation of the first line of result tables
               mat res_beta = indiv_beta       //Mean of the estimators
               mat res_sd = indiv_sd           //Standard deviation of the estimators
               mat res_IC_inf = indiv_IC_inf   //Lower bound of the confidence interval of the estimators
               mat res_IC_sup = indiv_IC_sup   //Upper bound of the confidence interval of the estimators
               mat res_puiss = indiv_puiss     //Estimators power
            }
            else {        //add of a line on result tables already existing
               mat res_beta = res_beta\indiv_beta
               mat res_sd = res_sd\indiv_sd
               mat res_IC_inf = res_IC_inf\indiv_IC_inf
               mat res_IC_sup = res_IC_sup\indiv_IC_sup
               mat res_puiss = res_puiss\indiv_puiss
            }
         }
      }

      /* Rename the name of the matrix column */
      matrix colnames res_beta = "Scenario" "ICC" "True" "Pooling" "SIMEX" "Reg-Cal"
      matrix colnames res_sd = "Scenario" "ICC" "True" "Pooling" "SIMEX" "Reg-Cal"
      matrix colnames res_IC_inf = "Scenario" "ICC" "Pooling" "SIMEX" "Reg-Cal"
      matrix colnames res_IC_sup = "Scenario" "ICC" "Pooling" "SIMEX" "Reg-Cal"
      matrix colnames res_puiss = "Scenario" "ICC" "True" "Pooling" "SIMEX" "Reg-Cal"

      /*-------------------------------------------------------------------------------------------------------*/
      /* MATRIX TRANSFORMATION IN VARIABLES + RENAME OF THESE VARIABLES */

      /* Mean of beta */
      svmat res_beta
      rename res_beta1 num_sce
      rename res_beta2 ICC
      rename res_beta3 beta_true
      rename res_beta4 beta_pool
      rename res_beta5 beta_simex
      rename res_beta6 beta_rcal

      /* Difference between beta mean and the true value of beta */
      qui gen perc_bias_pool = abs((beta_pool - beta_true) / beta_true) * 100
      qui gen perc_bias_simex = abs((beta_simex - beta_true) / beta_true)* 100
      qui gen perc_bias_rcal = abs((beta_rcal - beta_true) / beta_true) * 100

      /* Variance of beta */
      svmat res_sd
      drop res_sd1
      drop res_sd2
      rename res_sd3 sd_true
      rename res_sd4 sd_pool
      rename res_sd5 sd_simex
      rename res_sd6 sd_rcal

      /* Confidence interval */
         /* Lower bound */
      svmat res_IC_inf
      drop res_IC_inf1
      drop res_IC_inf2
      rename res_IC_inf3 IC_inf_pool
      rename res_IC_inf4 IC_inf_simex
      rename res_IC_inf5 IC_inf_rcal
         /* Upper bound */
      svmat res_IC_sup
      drop res_IC_sup1
      drop res_IC_sup2
      rename res_IC_sup3 IC_sup_pool
      rename res_IC_sup4 IC_sup_simex
      rename res_IC_sup5 IC_sup_rcal

      /* Power */
      svmat res_puiss
      drop res_puiss1
      drop res_puiss2
      rename res_puiss3 puiss_true
      rename res_puiss4 puiss_pool
      rename res_puiss5 puiss_simex
      rename res_puiss6 puiss_rcal

      /* Round of variables */
      global var_list = "pool simex rcal"

      qui replace beta_true = round(beta_true,1)

      foreach var in $var_list {
         qui replace beta_`var' = round(beta_`var',1)
         qui replace sd_`var' = round(sd_`var',0.1)
         qui replace IC_inf_`var' = round(IC_inf_`var',0.1)
```

```
436         qui replace IC_sup_`var' = round(IC_sup_`var',0.1)
437         qui replace puiss_`var' = round(puiss_`var',0.01)
438         qui replace perc_bias_`var' = round(perc_bias_`var',1)
439
440         qui tostring(IC_inf_`var'), gen(string_IC_inf_`var') force format(%7.0f)
441         qui tostring(IC_sup_`var'), gen(string_IC_sup_`var') force format(%7.0f)
442         qui gen IC_`var' = "[" + string_IC_inf_`var' + " ; " + string_IC_sup_`var' + "]" ///
443            if string_IC_inf_`var' != "." & string_IC_sup_`var' != "."
444      }
445
446      /* Back up of data set containing results */
447      save "${results}\res_beta${b1}_n${nb_indiv}", replace
448
449      /* Exportation of results on Excel */
450      export excel num_sce ICC beta_true beta_pool IC_pool puiss_pool perc_bias_pool ///
451         beta_simex IC_simex puiss_simex perc_bias_simex ///
452         beta_rcal IC_rcal puiss_rcal perc_bias_rcal ///
453         using "${results}\tab_res_beta${b1}_n${nb_indiv}.xls", firstrow(variables) replace
454
455
456      /*-------------------------------------------------------------------------------------------------------*/
457      /* GRAPHICS */
458
459      /* Percent bias */
460      graph bar perc_bias_pool perc_bias_simex perc_bias_rcal, over(ICC, relabel(1 "0.2" 2 "0.6")) ///
461         over(num_sce, relabel(1 "Scenario 1" 2 "Scenario 2" 3 "Scenario 3" 4 "Scenario 4")) bargap(-30) ///
462         bar(1, color(blue)) bar(2, color(red)) bar(3, color(green)) ///
463         ytitle("Bias in the effect estimate (%)", size(medlarge))  ///
464         legend(label(1 "Pooling method") label(2 "Simex") label(3 "Regression-Calibration") cols(3) span ///
465            bmargin(zero) ) name(graph_bias, replace)
466
467      /* Power */
468      graph bar (mean) puiss_pool puiss_simex puiss_rcal, over(ICC, relabel(1 "0.2" 2 "0.6")) ///
469         over(num_sce, relabel(1 "Scenario 1" 2 "Scenario 2" 3 "Scenario 3" 4 "Scenario 4")) bargap(-30) ///
470         bar(1, color(blue))  bar(2, color(red)) bar(3, color(green)) ///
471         exclude0 ylabel(0.4(0.1)1) ///
472         ytitle("Power", size(medlarge))  ///
473         legend(label(1 "Pooling method") label(2 "Simex") label(3 "Regression-Calibration") cols(3) span ///
474            bmargin(zero) ) name(graph_puiss, replace)
475
476      timer off 1
477      timer list
478
479      /*-------------------------------------------------------------------------------------------------------*/
480
481
482
```