Manuscript STOTEN-D-20-13879 Jedynak et al.

Response to the Reviewer

We are very grateful for the opportunity to submit a revised version of our manuscript entitled "Prenatal exposure to a wide range of environmental chemicals and child behaviour between 3 and 7 years of age - an exposome-based approach in 5 European cohorts" for publication in the Science of the Total Environment Journal. We appreciate the time and effort that the Reviewer dedicated to providing feedback on our manuscript and are grateful for the insightful comments and valuable improvements to our work. We have incorporated most of the suggestions made by the Reviewer. Please see below for a point-by-point response to the reviewer's comments and concerns.

Reviewer: The manuscript by Jedynak et al. addresses a relevant study question during an important window of development such as pregnancy in 5 European cohorts. Indeed, this is one of the few studies examining the role of chemical mixtures on children's behavior. The authors studied 708 mother-child pairs across Europe and found that higher prenatal urinary BPA and MnBP concentrations were associated with more externalizing problems, which is in line with both the epidemiological and toxicological evidence available. On the contrary, other exposures including Copper and persistent organic pollutants (POPs) showed inverse associations, which in the case of POPs could be influenced by adipose tissue kinetics. Overall, the study objective is very relevant, the design appropriate and the manuscript is concise and very well-written, so I feel this will be a nice contribution to the field. However, I believe the following comments will help to improve the manuscript:

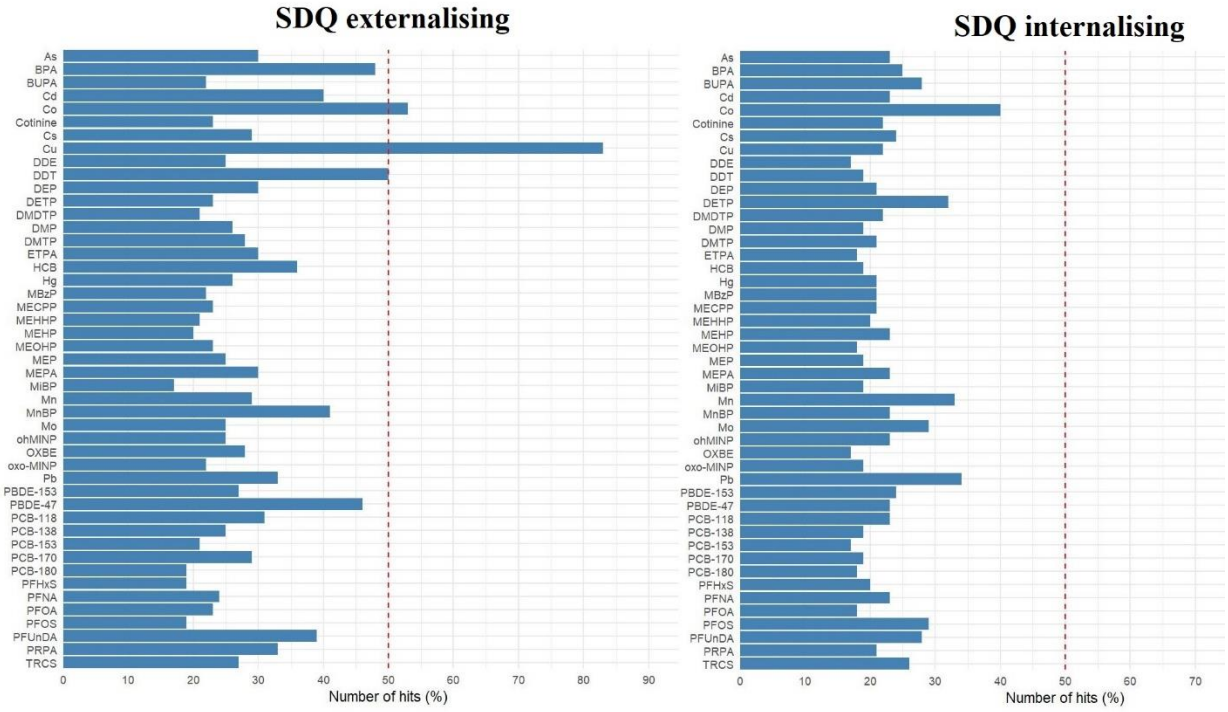
Major comments:

Reviewer's comment 1: In this study you decided to implement LASSO for variable selection. In a previous study you used Elastic-net (ENET) models (Philippat et al., 2019 doi: 10.1289/EHP3523). When comparing results from LASSO and ExWAS, it appears that LASSO was too restrictive. Would it be possible to additionally implement ENET and compare performance with LASSO?

Authors' Response: In this study, we did not implement the ENET procedure due to the lack of ready-to-use tools (neither in R or STATA) to apply ENET on multiply imputed data where the outcome variables (in our case SDQ scores) follow a negative binomial distribution. Obtaining the meaningful estimates for multiply imputed data directly from the penalized regression model would require a method to aggregate the coefficients of selected variables for each imputed dataset and to assign a standard error for the estimates of variables that were not selected. These operations are not straightforward and, to our best knowledge, no tools for performing them on data with negative binomial outcome exist. We did not

encounter such issues for the paper cited by the Reviewer (Philippat et al., 2019), since analyses were based on the non-imputed data and normally distributed outcome.

While, as described above, obtaining meaningful ENET coefficients is not straightforward when relying on multiple imputed datasets and non-gaussian outcome, ENET can still be used to select variables that can then be fitted into a new, non-penalized multiple regression model which produces the final estimates. For multiply imputed data this can be achieved by applying the same method as we used for LASSO, i.e. by fitting the ENET on each of the 100 imputed datasets and retaining an exposure only if it was selected in at least 50% of runs. Below we present the results of the variable selection using ENET, with the optimal α chosen by maximizing the prediction log-likelihood using 10-fold cross-validation and the corresponding λ+1SE. In line with the result obtained for the LASSO, the ENET algorithm did not retain any biomarker for the SDQ internalising score (Figure 1). For the externalising score, the ENET selected copper, a biomarker also selected by the LASSO, as well as two other biomarkers (cobalt and dichlorodiphenyltrichloroethane) that have not been identified by LASSO.



Revision Figure 1: Percentage of times each biomarker was selected by the ENET algorithm within 100 runs (one per imputed dataset). Red dashed line represents the threshold of 50% of runs. Models were adjusted for cohort, season of conception, child sex and age at SDQ assessment, parity, and maternal factors: education level, work status, age, pre- pregnancy BMI, and prenatal active smoking status.

Effect estimates obtained by fitting variables selected by ENET to a non-penalized multiple regression model are shown in Table 1 below. In this model, only copper was significantly associated with the externalising behavioural score and its estimates and CIs were almost identical as those obtained for the ExWAS. Since the results obtained with ENET were very similar to those for LASSO and ExWAS, we decided not to include this analysis in the manuscript.

Revision Table 1: Adjusted associations between prenatal exposure to environmental contaminants and SDQ externalising score selected by ENET (n = 708).

Behavioural outcome	Exposures included in the model	Main ExWAS		Multiple non-penalized regression model	
		IRR (95%CI)	p value	IRR (95%CI)	p value
SDQ externalising score	Cobalt (Co)	1.10 (0.98; 1.24)	0.112	1.11 (0.98; 1.25)	0.094
	Copper (Cu)	0.90 (0.82; 0.98)	0.021	0.90 (0.82; 0.99)	0.024
	Dichlorodiphenyltrichloroethane (DDT)	0.92 (0.84; 1.00)	0.045	0.93 (0.85; 1.01)	0.075

The multiple regression model was adjusted for cohort, season of conception, child sex and age at SDQ assessment, parity, and maternal factors: education level, work status, age, pre- pregnancy BMI, and prenatal active smoking status. Abbreviations: BMI = body mass index. CI = confidence interval of the IRR estimate. ExWAS = exposome-wide association study. IRR = incident rate ratio. SDQ = Strengths and Difficulties Questionnaire.

Reviewer's comment 2: Results section 3.2.2. Internalizing score. In ExWAS, the authors selected a threshold of $p<0.05$ to select associations. While I understand this approach, the authors face one ironical finding in their analysis: For MnBP-externalizing, the p-value was 0.048, and the association was selected. For DETP-internalizing, the p-value was 0.053, and this association has not even mentioned in the manuscript. If one checks if this potential association is driven by a unique cohort, then finds that there is a consistent pattern towards more internalizing problems in four out of the five cohorts examined: BiB 2.04 (1.33;3.12); EDEN 1.12 (0.91;1.36); INMA 0.98 (0.81;1.18); KANC 1.15 (0.78;1.67) and RHEA 1.11 (0.92;1.33). Therefore, I think the DETP association should be acknowledged in both the results section and the abstract, and highlight that future studies need to confirm or ruled out this association.

Authors' Response: As suggested by the Reviewer, the association between DETP and SDQ internalizing score is now mentioned in the abstract, results, discussion and Table 2.

Reviewer's comment 3: Following the previous example, I think the authors could maintain the p-threshold at <0.05, but also add that the direction and pattern of associations across the five cohorts was also accounted in the selection of the most relevant associations. Thus, this study identified patterns of

associations that, although not significant at the conventional level, would be worth it to highlight as candidate associations requiring further confirmation. Examples are: cadmium and cobalt for externalizing problems; and cobalt, manganese, BUPA and TRCS for internalizing problems.

Authors' Response: We agree that the consistency of the results across the cohorts is an important criterion to select the most relevant associations between prenatal exposures and SDQ scores. However, we do not find it feasible to rely on the direction and pattern of association across the cohorts to select the associations to be mentioned in the result section. Effect estimates obtained for each cohort for the biomarkers mentioned by the Reviewer had wide confidence intervals that most often included 1, which made it difficult to define the direction and pattern of those associations (see Table 2 below). Moreover, apart from the mentioned issues, the association between Co and the SDQ internalising score mentioned by the Reviewer showed a moderate heterogeneity across the cohorts ($I^2 = 0.42$) pointing rather towards the inconsistency of the results rather than any common pattern. For these reasons we decided to not expand the list of associations mentioned in the results section.

Revision Table 2: Adjusted associations between prenatal exposure to environmental contaminants suggested by the Reviewer and SDQ externalising (upper panel) and internalising scores (lower panel), extracted from the Appendix Table 6 and 7 (n = 708).

Exposure	ExWAS		ExWAS with cohort-exposure interaction[a]					I^2
			IRR (95%CI)					
	IRR (95%CI)	p value	BiB	EDEN	INMA	KANC	RHEA	
Cadmium (Cd)	1.08 (0.97; 1.20)	0.182	1.17 (0.76;1.80)	1.11 (0.92;1.33)	1.03 (0.88;1.22)	1.09 (0.92;1.29)	1.12 (0.97;1.29)	<0.001
Cobalt (Co)	1.10 (0.98; 1.24)	0.112	1.21 (0.85;1.73)	1.21 (0.99;1.47)	1.04 (0.86;1.26)	0.99 (0.75;1.30)	1.09 (0.92;1.30)	<0.001

Exposure	ExWAS		ExWAS with cohort-exposure interaction[a]					I^2
			IRR (95%CI)					
	IRR (95%CI)	p value	BiB	EDEN	INMA	KANC	RHEA	
Cobalt (Co)	1.10 (0.96; 1.27)	0.179	1.50 (1.02;2.21)	1.06 (0.85;1.33)	1.08 (0.84;1.38)	0.77 (0.54;1.09)	1.23 (0.98;1.55)	0.420
Manganese (Mn)	1.11 (0.98; 1.26)	0.099	1.07 (0.70;1.63)	1.07 (0.87;1.32)	1.12 (0.87;1.44)	1.03 (0.76;1.40)	1.19 (0.98;1.44)	<0.001
N-butyl paraben (BUPA)	1.09 (0.98; 1.21)	0.121	1.24 (0.89;1.72)	1.03 (0.83;1.27)	1.24 (1.02;1.51)	0.99 (0.67;1.46)	1.00 (0.82;1.21)	0.111
Triclosan (TRCS)	1.08 (0.96; 1.21)	0.204	1.83 (1.15;2.90)	1.12 (0.91;1.39)	1.00 (0.81;1.25)	1.04 (0.69;1.55)	1.02 (0.82;1.25)	<0.001

The regression models were adjusted for cohort, season of conception, child sex and age at SDQ assessment, parity, and maternal factors: education level, work status, age, pre- pregnancy BMI, and prenatal active smoking status. Abbreviations: BMI = body mass index. CI = confidence interval of the IRR estimate. ExWAS = exposome-wide association study. IRR = incident rate ratio. SDQ = Strengths and Difficulties Questionnaire. BiB = Born in Bradford. EDEN = Étude des Déterminants Pré et Postnatals du Développement et de la Santé de l'Enfant. INMA = Infancia y Medio Ambiente. KANC = Kaunas Cohort. RHEA = Mother-Child Cohort in Crete.

Reviewer's comment 4: Discussion lines 274-277: Good point with the residual confounding advantage of pooling the five cohorts. Just after this sentence, or in any other part, you should also acknowledge that the variability among cohorts and procedures also presents limitations that would probably tend to introduce some random noise and bias toward the null. Could you briefly elaborate this part? For example, if BPA or MnBP was measured by different labs, without an interlaboratory comparison, this will tend to increase variability in exposure-SDQ associations. If urine samples were collected at different trimesters in each cohort, this will also add variability due to normal dilution effects, etc.

Authors' Response: The Reviewer is right. Since the cohorts were recruited before the start of the HELIX project, collection of biological samples during pregnancy was not harmonized leading to different timings (i.e., different trimester) for exposure assessment across cohorts. It is also true that, for some cohorts, the same biomarker was sometimes assessed in different laboratories. However, the results of interlaboratory comparisons performed in the framework of the HELIX protocol show a high correlation between assessments performed by different laboratories (see Table 3 below). For instance, phenol metabolites were assessed by both the Norwegian Institute of Public Health (NIPH) and the Centers for Disease Control and Prevention (CDC) in 35 urine samples of the EDEN cohort. Correlation coefficients between these two measurements were high (≥ 0.90, see Table 3 below). Results for all interlaboratory comparisons performed within the Helix framework can be found in the Supplementary material of (Tamayo-Uria et al. 2019), pp. 16-19.

Revision Table 3: Spearman correlation coefficients between 12 phenol urinary maternal samples analysed in the NIPH and CDC laboratories.

	NIPH						
CDC	MEPA	ETPA	PRPA	BUPA	BPA	OXBE	TRCS
MEPA	1.0	-	-	-	-	-	-
ETPA	-	0.98	-	-	-	-	-
PRPA	-	-	0.90	-	-	-	-
BUPA	-	-	-	1.0	-	-	-
BPA	-	-	-	-	0.90	-	-
OXBE	-	-	-	-	-	0.91	-
TRCS	-	-	-	-	-	-	0.99

Abbreviations: BPA = bisphenol A. BUPA = n-butyl paraben. ETPA = ethyl paraben. MEPA = methyl paraben. OXBE = oxybenzone. PRPA = propyl paraben. TRCS = triclosan. CDC = Centers for Disease Control and Prevention. NIPH = Norwegian Institute of Public Health.

All these issues are now mentioned in the discussion section: "*On the other hand, since the cohorts were recruited before the start of the HELIX project, collection of biological samples during pregnancy was not harmonized leading to different timings (i.e., different trimester) for exposure assessment across*

cohorts. Additionally, for some cohorts the same exposure biomarker was sometimes assessed by different laboratories (see Appendix Table 2), which may partly explain the between-cohort heterogeneity of the results observed for some exposures. This should not have a strong impact on our results since the interlaboratory comparisons performed in the framework of the HELIX protocol suggested a high correlation between assessments performed in different laboratories. For instance, correlation coefficients between phenol urinary concentrations measured by the Norwegian Institute of Public Health and Centers for Disease Control and Prevention in 12 maternal samples of the EDEN cohort were ≥ 0.90 (Supplementary of (Tamayo-Uria et al. 2019)).".

General comments:

Reviewer's comment: Please check the manuscript for blank spaces throughout the text.

Authors' Response: We have revised the manuscript for blank spaces. As they are not present in the .docx document, we assume these were artefacts of PDF conversion. We will make sure they will not appear in the revised version of the manuscript.

Reviewer's comment: Check Appendix Tables 5 and 6, and complete estimates, so all appear with two decimals, even if it is for example 1.00 instead of 1, or 0.90 instead of 0.9, to be consistent throughout the table.

Authors' Response: We have corrected the tables accordingly.

Reviewer's comment: Highlights: Copper and POPs findings are not mentioned.

Authors' Response: We have mentioned the association between copper and SDQ score in the highlights: "*Copper was associated with decreased behavioural scores*". However, we decided to not mention the POPs findings because they were mainly seen among women with insufficient weight gain during pregnancy and were not consistent with the literature.

Reviewer's comment: Abstract: Line 22 "The association with PFOS" instead of "An".

Authors' Response: We have corrected the sentence.

Reviewer's comment: Introduction: I like the introduction, concise and informative.

Authors' Response: Thank you.

Reviewer's comment: Line 69: In a previous STOTEN paper, Kim et al., analyzed different chemical families in relation to child neurodevelopment:

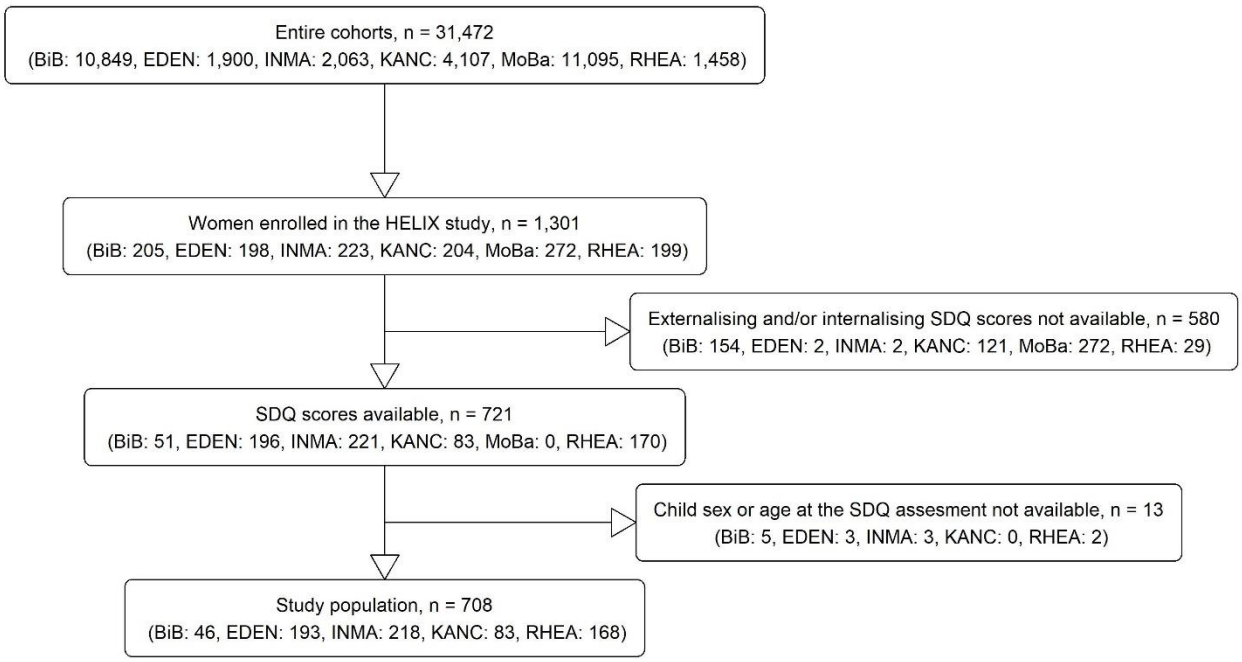
Kim et al., 2018. Association between maternal exposure to major phthalates, heavy metals, and persistent organic pollutants, and the neurodevelopmental performances of their children at 1 to 2years of age- CHECK cohort study. Sci Total Environ. doi: 10.1016/j.scitotenv.2017.12.058.

Authors' Response: We have updated the reference.

Methods

Reviewer's comment: 2.1 Study design and population: A flow-chart explaining the inclusion of mother-child couples from each cohort would be helpful.

Authors' Response: We have added a flow-chart to the appendix (Appendix Figure 1). See also below.



Appendix Figure 1: Study flow-chart. Abbreviations: BiB = Born in Bradford. EDEN = Étude des Déterminants Pré et Postnatals du Développement et de la Santé de l'Enfant. INMA = Infancia y Medio Ambiente. KANC = Kaunas Cohort. MoBa = Norwegian Mother, Father and Child Cohort Study. RHEA = Mother-Child Cohort in Crete. SDQ = Strengths and Difficulties Questionnaire. Adapted from (Maitre et al. 2018).

Reviewer's comment: Line 85: Strenght[s] and Difficulties Questionnaire [delete "s"].

Authors' Response: We have corrected the sentence.

Reviewer's comment: Lines 84-85: you mention child twice in the same sentence.

Authors' Response: We have removed the repeated word.

Reviewer's comment: 2.2 Exposure assessment: Many different chemicals were measured. In lines 96 and 97 the authors cite two papers, but this seems insufficient to described the methodologies, laboratories and quality controls followed for all chemicals and metabolites. Was each chemical family assessed by the same lab in all the cohorts? Interlaboratory comparisons were available? In the supplementary material, a table providing brief but relevant information for each chemical compound would be helpful. For example, it would be nice to know the laboratory in which the sample was analyzed in each cohort, the method (HPLC, GCMS…), the exact biological sample used in each cohort (e.g. for PFAS, while serum/plasma can be compared, PFAS measured in whole blood would probably lead to a reduction in PFAS concentrations). This information can help to decide to which extent cohorts and measurements are comparable, and to perform a more informative interpretation of study findings in light of the variability observed among cohorts.

Authors' Response: We have added a table with the requested information to the appendix (Appendix Table 2). See also below.

Appendix Table 1: Exposure assessment – biological matrices, timing of sample collection, laboratories, methodologies and quality controls (adapted from (Haug et al. 2018)).

Family of exposures	Cohort BiB	EDEN	KANC	INMA	RHEA
OCs					
Biological matrix	Serum/plasma	Serum	-	Serum	Serum
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	13.7 (2.0)	39.4 (1.3)	14.1 (3.7)
Laboratory	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	LSPG (Spain)	National Institute for Health and Welfare, Chemical Exposure Unit, Kuopio (Finland)
Analytical method	GC-MS/MS	GC-MS/MS	GC-MS/MS	GC-MS	GC-MS/MS
Interlaboratory comparison	We participated 3 times in the AMAP interlaboratory comparison study during the period when HELIX samples were analysed. Each round included 3 samples. The samples are spiked, and most analytes are in a much higher concentration range than our samples usually are. Despite these factors Z- scores below 2 were obtained for all contaminants except for PCB-153, and for PCB-170 in 1 sample.				
Standard reference material	2 different SRMs were obtained from NIST, Organic Contaminants in Human Serum, non-fortified (1957) and fortified (1958). The RSDs of the 6 different injections of the non-fortified sample (1957) were 2-18%. The deviation from the certified values was 2-20%, except for HCB (50%) and PCB-153 (76%), where a small blank contribution may be the problem. The RSDs of the 6 different injections of the fortified sample (1958) were 2-20%, and for most analytes around 5%. The deviation from the certified values were 2-17%, except for DDT (-40%).				
Reference for the analytic method	Modified (Caspersen et al. 2016)	Modified (Caspersen et al. 2016)	Modified (Caspersen et al. 2016)	(Goñi et al. 2007)	(Koponen et al. 2013)
PBDEs					
Biological matrix	Serum/plasma	Serum	-	Serum	Serum
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	13.7 (2.0)	39.4 (1.3)	14.1 (3.7)
Laboratory	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	National Institute for Health and Welfare, Chemical Exposure Unit, Kuopio (Finland)
Analytical method	GC-MS/MS	GC-MS/MS	GC-MS/MS	GC-MS	GC-MS/MS
Interlaboratory comparison	We participated 3 times in the AMAP interlaboratory comparison study during the period when HELIX samples were analysed. Each round included 3 samples. The samples are spiked, and most analytes are in a much higher concentration range than our samples usually are. Despite these factors Z- scores below 2 were obtained for all contaminants.				
Standard reference material	2 different SRMs were obtained from NIST, Organic Contaminants in Human Serum, non-fortified (1957) and fortified (1958). The RSDs of the 6 different injections of the non-fortified sample (1957) were 2-18%. The deviation from the certified values was 2-20%. The RSDs of the 6 different injections of the fortified sample (1958) were 2-20%, and for most analytes around 5%. The deviation from the certified values were 2-17%.				
Reference for the analytic method	Modified (Caspersen et al. 2016)	Modified (Caspersen et al. 2016)	Modified (Caspersen et al. 2016)	Modified (Caspersen et al. 2016)	(Koponen et al. 2013)
PFASs					
Biological matrix	Serum/plasma	Serum	Whole blood	Plasma	Serum
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	13.7 (2.0)	39.4 (1.3)	14.1 (3.7)

Laboratory	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	Institute for Occupational Medicine, RWTH Aachen University (Germany)	NIPH (Norway)
Analytical method	GC-MS	GC-MS	GC-MS	Online column-switching LC–MS–MS analysis	GC-MS
Interlaboratory comparison	We participated 3 times in the AMAP interlaboratory comparison study during the period where HELIX samples were analysed. Each round included 3 samples. The samples are spiked in a wide concentration range. Z- scores below 2 were obtained for all contaminants except for PFUnDA in 2 samples.				
Standard reference material	In total 15 samples from 4 rounds of the AMAP interlaboratory comparison study were analysed during the period where HELIX samples were analysed, and the mean deviation from the assigned value varied between 8 and 17%.				
Reference for the analytic method	(Haug et al. 2009)	(Haug et al. 2009)	(Poothong et al. 2017)	(Manzano-Salgado et al. 2015)	(Haug et al. 2009)
Metals and semi-metals					
Biological matrix	Whole blood	Whole blood	Whole blood	Cord whole blood	Whole blood
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	13.7 (2.0)	39.4 (1.3)	14.1 (3.7)
Laboratory	ALS Scandinavia (Sweden)	ALS Scandinavia (Sweden)	ALS Scandinavia (Sweden)	Hg; LSPA (Spain)	ALS Scandinavia (Sweden)
Analytical method	ICP-SFMS	ICP-SFMS	ICP-SFMS	AAS	ICP-SFMS
Standard reference material	In total 15 samples of 2 Seronorm whole blood reference materials were provided to ALS Scandinavia which performed analyses of toxic and essential elements of the HELIX samples. The samples were blinded for the laboratory. The mean deviation from the assigned value varied between 1 and 17%, except for Co which was 53%.				
Reference for the analytic method	(Rodushkin et al. 2000)	(Rodushkin et al. 2000)	(Rodushkin et al. 2000)	(Ramon et al. 2011)	(Rodushkin et al. 2000)
Phthalate metabolites					
Biological matrix	Urine	Urine	-	Urine	Urine
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	34.2 (1.3)	NA	14.1 (3.7)
Laboratory	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	Bioanalysis Research Group at the Hospital del Mar Medical Research Institute (Spain)	NIPH (Norway)
Analytical method	LC-MS/MS	LC-MS/MS	LC-MS/MS	HPLC-MS	LC-MS/MS
Standard reference material	In total 42 samples of NIST reference material SRM 3673 were analysed during the period where HELIX samples were analysed, and the mean deviation from the assigned value varied between 1 and 25%.				
Reference for the analytic method	(Sabaredzovic et al. 2015)	(Sabaredzovic et al. 2015)	(Sabaredzovic et al. 2015)	(Valvi et al. 2015)	(Sabaredzovic et al. 2015)
Phenols					
Biological matrix	Urine	Urine	-	Urine	Urine
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	34.2 (1.3)	NA	14.1 (3.7)
Laboratory	NIPH (Norway)	National Center for Environmental Health laboratory at the CDC (US)	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)
Analytical method	UPLC-MS/MS	Described in (Philippat et al. 2012)	UPLC-MS/MS	UPLC-MS/MS	UPLC-MS/MS

Interlaboratory comparison	We participated twice in the External Quality Assessment Scheme for organic substances in urine for BPA and TRCS. Each round included 2 samples. Z- scores below 2 were obtained for both contaminants in all samples.				
Standard reference material	In total 145 samples of NIST reference material SRM 3673 were analysed during the period where HELIX samples were analysed, and the mean deviation from the assigned value varied between 1 and 19%, except BPA which was 63%.				
Reference for the analytic method	(Sakhi et al. 2018)	(Philippat et al. 2012)	(Sakhi et al. 2018)	(Sakhi et al. 2018)	(Sakhi et al. 2018)
OP pesticide metabolites					
Biological matrix	Urine	Urine	-	Urine	Urine
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	34.2 (1.3)	NA	14.1 (3.7)
Laboratory	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)	NIPH (Norway)
Analytical method	UPLC-TOF	UPLC-TOF	UPLC-TOF	UPLC-TOF	UPLC-TOF
Reference	(Cequier et al. 2016)	(Cequier et al. 2016)	(Cequier et al. 2016)	(Cequier et al. 2016)	(Cequier et al. 2016)
Cotinine					
Biological matrix	Urine	Urine	-	Urine	Urine
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	34.2 (1.3)	NA	14.1 (3.7)
Laboratory	Fürst Medical Analysis Laboratory (Norway)	Fürst Medical Analysis Laboratory (Norway)	Fürst Medical Analysis Laboratory (Norway)	Public Health Laboratory of Bilbao - LSPPV (Spain)	Fürst Medical Analysis Laboratory (Norway)
Analytical method	The Immulite® 2000 Nicotine Metabolite (Cotinine) 600 Test on an Immulite 2000 XPi from Siemens Healthineers	The Immulite® 2000 Nicotine Metabolite (Cotinine) 600 Test on an Immulite 2000 XPi from Siemens Healthineers	The Immulite® 2000 Nicotine Metabolite (Cotinine) 600 Test on an Immulite 2000 XPi from Siemens Healthineers	LC-MS	The Immulite® 2000 Nicotine Metabolite (Cotinine) 600 Test on an Immulite 2000 XPi from Siemens Healthineers
Reference for the analytic method	-	-	-	(Aurrekoetxea et al. 2013)	-
Creatinine					
Biological matrix	Urine	Urine	-	Urine	Urine
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	34.2 (1.3)	NA	14.1 (3.7)
Laboratory	Fürst Medical Analysis Laboratory (Norway)	National Center for Environmental Health laboratory at the CDC (US)	Fürst Medical Analysis Laboratory (Norway)	Public Health Echevarne Laboratory of Barcelona (Spain)	Fürst Medical Analysis Laboratory (Norway)
Analytical method	AU680 Chemistry System form Beckman Coulter using DRI® Creatinine-Detect® Test	Enzymatic reaction using a Roche Hitachi 912 chemistry analyzer (Roche Hitachi, Basel, Switzerland)	AU680 Chemistry System form Beckman Coulter using DRI® Creatinine-Detect® Test	Jaffé method - Beckman Coulter© AU5400	AU680 Chemistry System form Beckman Coulter using DRI® Creatinine-Detect® Test
Reference for the analytic method	-	-	-	-	-
Lipids					
Biological matrix	Serum/plasma	Serum	-	Serum	Serum
Timing of sample collection, mean GW (SD)	26.6 (1.4)	26.1 (1.2)	13.7 (2.0)	39.4 (1.3)	14.1 (3.7)

Laboratory	Fürst Medical Analysis Laboratory (Norway)	Fürst Medical Analysis Laboratory (Norway)	Fürst Medical Analysis Laboratory (Norway)	Laboratorio de Salud Pública de Bizkaia (Spain)	Medicon Hellas SA, Gerakas (Greece)
Analytical method	ADVIA® Chemistry XPT System, and the FS kit from DiaSys was used to measure concentrations of phospholipids	ADVIA® Chemistry XPT System, and the FS kit from DiaSys was used to measure concentrations of phospholipids	ADVIA® Chemistry XPT System, and the FS kit from DiaSys was used to measure concentrations of phospholipids	Cobas Mira self-analyzer (Roche Diagnostic, Basel, Switzerland) using an Enzymatic-Colorimetric method with Spinreact reagents	Standard enzymatic method
Reference for the analytic method	-	-	-	-	-

Abbreviations: GW = gestational week. NA = not available. SD = standard deviation. BPA = bisphenol A. DDT = dichlorodiphenyltrichloroethane. HCB = hexachlorobenzene. OC = organochlorine compound. OP = organophosphate. PBDE = polybrominated diphenyl ether. PCB = polychlorinated biphenyl. PFAS = per- and polyfluoroalkyl substance. PFUnDA = perfluoroundecanoate. TRCS = triclosan. RSD = relative standard deviation. SRM = Standard reference material. AAS = thermal decomposition, amalgamation and atomic absorption spectrometry. GC-MS/MS = gas chromatography coupled with tandem mass spectrometry. GC-MS = gas chromatography mass spectrometry. GC-MS-NICI = gas chromatography-negative-ion chemical ionization mass spectrometry. HPLC-MS = ultra-performance liquid chromatography coupled to tandem mass spectrometry. ICP-SFMS = conductively coupled plasma-sector field mass spectrometry. LC-MS/MS = liquid chromatography-mass spectrometry. Q-ICP-MS = inductively coupled plasma quadruple mass spectrometry. SPE-HPLC–MS/MS = on-line solid-phase extraction coupled to isotope dilution high performance liquid chromatography-tandem mass spectrometry. UPLC-MS/MS = ultra-performance liquid chromatography-tandem mass spectrometry. UPLC-TOF = ultra-performance liquid chromatography coupled to time-of-flight mass spectrometry. AMAP = Arctic Monitoring and Assessment Program “Ring Test for Persistent Organic Pollutants in Human Serum”. CDC = Centers for Disease Control and Prevention. LSPA = Laboratorio de Salud Pública de Alava. LSPG = Laboratorio de Salud Pública de Guipúzcoa. NIPH = Norwegian Institute of Public Health. NIST = National Institute of Standards and Technology, USA. BiB = Born in Bradford. EDEN = Étude des Déterminants Pré et Postnatals du Développement et de la Santé de l’Enfant. INMA = Infancia y Medio Ambiente. KANC = Kaunas Cohort. RHEA = Mother-Child Cohort in Crete.

2.3 Behavioral outcomes

Reviewer's comment: Lines 101-105: Apart from this rationale, it should be included that this approach was also selected to reduce the number of multiple comparisons performed. Subscales also provide interesting and predictive information, but in this case, investigating 47 exposures and all subscales, would lead to a multiple comparison issue difficult to disentangle, and a huge amount of data difficult to present to the reader.

Authors' Response: We have updated the text accordingly: *"Moreover, given our limited sample size and the large number of studied exposures, combining the SDQ sub-scales limited the number of performed tests."*

2.4. Statistical analysis

Reviewer's comment: Lines 107-108: reference for this sentence?

Authors' Response: We have added a relevant reference.

Reviewer's comment: Lines 109-110: "haemal lipophilic biomarker concentrations by total fat percentage" It reads as if you normalized POPs by total body fat percentage. Probably you want to say that lipophilic biomarker concentrations were standardized by serum lipids and expressed as ng/g of lipid.

Authors' Response: We clarified: "*Haemal lipophilic biomarker concentrations were standardized and expressed in ng/g of total lipids in serum or plasma.*"

Reviewer's comment: Why season of conception was included in the models? For specific biomarkers such as BP-3, season of sample collection would be relevant, but I do not totally see its role as a relevant confounder of the associations tested.

Authors' Response: Exposure to several environmental contaminants considered in our study is dependent on behaviours regulated by season, e.g. on time spent outdoors. That included exposure to some heavy metals (e.g. cadmium, lead and mercury that are common air pollutants (Tchounwou et al. 2012)) or substances present in sunscreens (e.g. oxybenzone also known as BP-3 (American Association of Pharmaceutical Scientists 2018)) or in other seasonal cosmetics (e.g. parabens). On the other hand, there is evidence on season of birth being associated with emotional and behavioural regulation in children (Asano et al. 2016) or risk of neurodevelopmental disorders (e.g. (Zerbo et al. 2011). For these reasons we decided to include the season of conception in the models.

Results

Reviewer's comment: Lines 162-165: Median externalizing score for BiB children was 0.5, in striking contrast with the rest of cohorts, in which the median was between 5 and 6. Based on this huge difference, what is your criteria to not exclude this cohort from the analysis at least in a sensitivity?

Authors' Response: We have provided an additional sensitivity analysis for the SDQ externalizing scores where we excluded the BiB cohort and compared the results with the main ExWAS. Results of this additional analysis are now included in the main text and in Table 2, see also below. Exclusion of the BiB cohort did not strongly affect our results for the SDQ externalising score, except of widening of the confidence interval and increase of p value for DDT.

Revision Table 4: Adjusted associations between the prenatal exposure to environmental contaminants and SDQ externalising scores, extracted from the Table 2.

Behavioural outcome	Exposure	Exposure family	ExWAS			ExWAS after exclusion of the BiB cohort		
			IRR (95%CI)	p value	FWER p value	IRR (95%CI)	p value	n
SDQ externalising score	BPA	Phenol	1.06 (1.01; 1.12)	0.028	0.842	1.06 (1.01; 1.12)	0.026	662
	Cu	Essential element	0.90 (0.82; 0.98)	0.021	0.631	0.91 (0.83; 1.00)	0.042	662
	DDT	Organochlorine compound	0.92 (0.84; 1.00)	0.045	1	0.94 (0.86; 1.03)	0.174	662
	MnBP	Phthalate	1.06 (1.00; 1.13)	0.048	1	1.06 (1.00; 1.13)	0.041	662
	PCB-138	Organochlorine compound	0.88 (0.79; 0.99)	0.035	1	0.88 (0.79; 0.99)	0.031	662
	PFUnDA	Perfluoroalkyl substance	0.92 (0.84; 0.99)	0.034	1	0.90 (0.82; 0.98)	0.013	662

The regression model was adjusted for cohort, season of conception, child sex and age at SDQ assessment, parity, and maternal factors: education level, work status, age, pre- pregnancy BMI, and prenatal active smoking status. Abbreviations: BiB = Born in Bradford. BMI = body mass index. CI = confidence interval of the IRR estimate. ExWAS = exposome-wide association study. FWER = family wise error rate adjustment. IRR = incident rate ratio. SDQ = Strengths and Difficulties. BPA = bisphenol A. Cu = copper. DDT = dichlorodiphenyltrichloroethane. MnBP = mono-n-butyl phthalate. PCB = polychlorinated biphenyl. PFUnDA = perfluoroundecanoate.

Reviewer's comment: Lines 195-197: Appendix Figure 1 is informative. It could be included in the main manuscript. In that case change "Addequate to Adequate".

Authors' Response: The figure was moved to the main manuscript with corrected legend.

Reviewer's comment: Table 1: Maternal work status. Numbers in the "overall distribution" column are interchanged between "unemployed" and "employed".

Authors' Response: We have corrected the numbers accordingly.

Reviewer's comment: Figure 1 footnote: Could you include the sample size for each of the cohorts? Or in any other place?

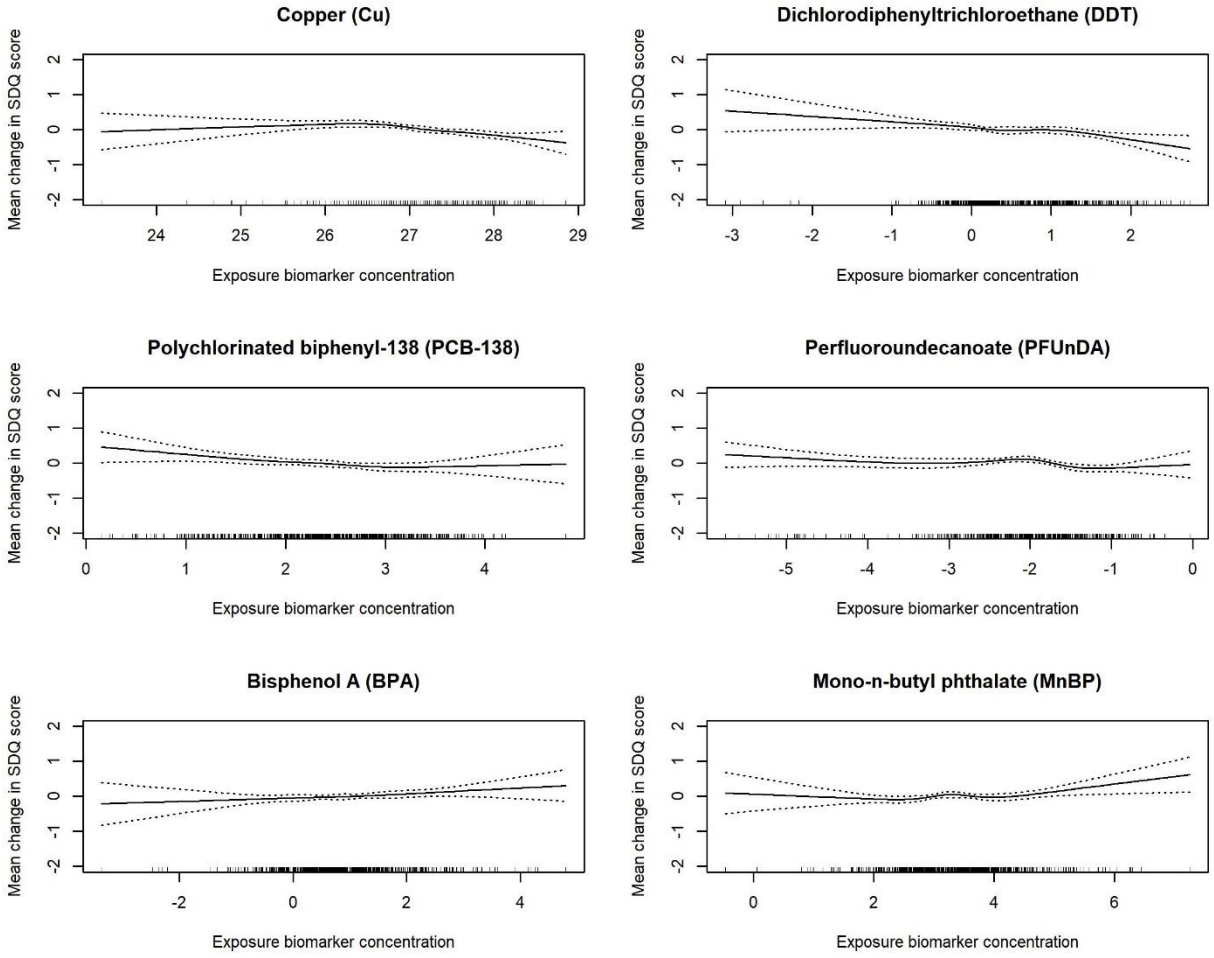
Authors' Response: We have added the requested information in the Figure 1 footnote.

Discussion

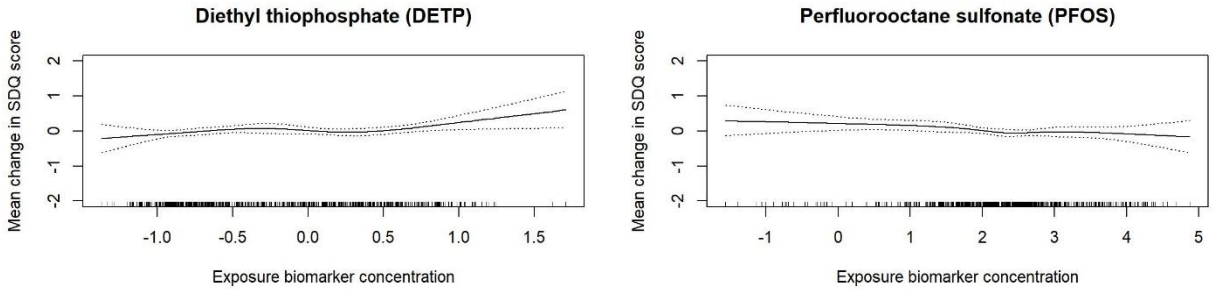
Reviewer's comment: Lines 213-214: Have you checked the functional shape of the Cu association using for example LOWESS or GAMs? For essential metals, as you have stated, too much or too little would cause problems. Thus, a monotonic dose-response is not expected.

Authors' Response: We now use the GAM with the restricted cubic splines function fitted on the biomarker concentrations that were detected in the LASSO and ExWAS as associated with the SDQ scores. Results are presented in the Appendix Figure 2, see also below. All detected exposure-SDQ score associations were linear.

(A)



(B)



Appendix Figure 2: GAMs with restricted cubic splines function fitted on the log_2 and IQR transformed prenatal concentrations of exposures selected by the LASSO and ExWAS as associated with the externalising (A) and internalising (B) SDQ scores. Solid line represents the fit and dashed line the standard error of the fit. The model was adjusted for cohort, season of conception, child sex and age at SDQ assessment, parity, and maternal factors: education level, work status, age, pre- pregnancy BMI, and prenatal active smoking status. Abbreviations: GAM = generalized additive model. ExWAS = exposome-wide association study. IQR = inter-quartile range. LASSO = least absolute shrinkage and selection operator. SDQ = Strengths and Difficulties Questionnaire.

Reviewer's comment: Lines 225-239: I agree with this paragraph. However, I think the authors should first mention that prenatal BPA has been associated with both externalizing and internalizing behaviors in previous studies (updated review Table 1 Mustieles and Fernández, 2020). Indeed, results from this study support that BPA may be related to both internalizing and externalizing problems, although externalizing problems were more evident. In Appendix Table 6, for internalizing problems, BPA showed the following overall estimate: 1.04 (0.98, 1.12) p=0.209. Additionally, four out of five cohorts showed positive estimators: BiB 1.14 (0.83;1.55); EDEN 1.16 (0.98;1.36); INMA 1.02 (0.93;1.13); KANC 0.98 (0.78;1.23) and RHEA 1.04 (0.92;1.19). Given that the BPA-behavior relationship is one of the most studied, it would be worth it to mention this trend.

Authors' Response: In the discussion section we have now acknowledged the previous studies that had reported associations with the SDQ internalizing score and we have mentioned that the effect estimate for this association was also positive in our study population: "*Previous studies also reported higher scores on the internalizing behaviour sub-scale in association with the prenatal exposure to bisphenol A* (Braun et al. 2011, 2017; Evans et al. 2014; Grohs et al. 2019; Harley et al. 2013; Li et al. 2020; Perera et al. 2012, 2016; Philippat et al. 2017; Roen et al. 2015). *While not significant (p value = 0.21), effect estimate for our study population also suggested a positive association between BPA and internalizing SDQ score (IRR: 1.04 (95%CI: 0.98, 1.12)).*"

Reviewer's comment: Lines 237-239: The experimental and epidemiologic data for BPA and behavior has been updated in Mustieles and Fernández 2020. Additionally, you can also cite Nesan et al., 2018 for

experimental studies.

Mustieles V, Fernández MF. Bisphenol A shapes children's brain and behavior: towards an integrated neurotoxicity assessment including human data. Environ Health. 2020 Jun 9;19(1):66. doi: 10.1186/s12940-020-00620-y.

Nesan D, Sewell LC, Kurrasch DM. Opening the black box of endocrine disruption of brain development: Lessons from the characterization of Bisphenol A. Horm Behav. 2018 May;101:50-58. doi: 10.1016/j.yhbeh.2017.12.001.

Authors' Response: We have updated the references according to the Reviewer's request.

Conclusions

Reviewer's comment: Lines 307-308: "These negative associations must be interpreted with caution as they were not reported in previous studies, [and could even result from changes in body composition during pregnancy]."

Authors' Response: We have corrected the sentence accordingly: "*Cu, DDT, PCB-138, PFOS and PFUnDA were negatively associated with child behaviour scores, suggesting lower risk of behavioural problems. These negative associations were not reported in previous studies and for lipophilic compounds (DDT and PCB-138) could even result from changes in body composition during pregnancy*."

Revision references:

American Association of Pharmaceutical Scientists. 2018. Clinical and environmental factors impact absorption of common sunscreen ingredient.

Asano R, Tsuchiya KJ, Harada T, Kugizaki Y, Nakahara R, Nakayasu C, et al. 2016. Season of birth predicts emotional and behavioral regulation in 18-month-old infants: Hamamatsu Birth Cohort for Mothers and Children (HBC Study). Front Public Health 4; doi:10.3389/fpubh.2016.00152. PMID: 27504441.

Aurrekoetxea JJ, Murcia M, Rebagliato M, López MJ, Castilla AM, Santa-Marina L, et al. 2013. Determinants of self-reported smoking and misclassification during pregnancy, and analysis of optimal cut-off points for urinary cotinine: a cross-sectional study. BMJ Open 3:e002034; doi:10.1136/bmjopen-2012-002034. PMID: 23355667.

Braun JM, Kalkbrenner AE, Calafat AM, Yolton K, Ye X, Dietrich KN, et al. 2011. Impact of early-life bisphenol A exposure on behavior and executive function in children. Pediatrics 128:873–882; doi:10.1542/peds.2011-1335. PMID: 22025598.

Braun JM, Muckle G, Arbuckle T, Bouchard MF, Fraser WD, Ouellet E, et al. 2017. Associations of prenatal urinary bisphenol A concentrations with child behaviors and cognitive abilities. Environ Health Perspect 125; doi:10.1289/EHP984. PMID: 28657891.

Caspersen IH, Kvalem HE, Haugen M, Brantsæter AL, Meltzer HM, Alexander J, et al. 2016. Determinants of plasma PCB, brominated flame retardants, and organochlorine pesticides in pregnant women and 3 year old children in The Norwegian Mother and Child Cohort Study. Environ Res 146:136–144; doi:10.1016/j.envres.2015.12.020. PMID: 26749444.

Cequier E, Sakhi AK, Haug LS, Thomsen C. 2016. Development of an ion-pair liquid chromatography-high resolution mass spectrometry method for determination of organophosphate pesticide metabolites in large-scale biomonitoring studies. J Chromatogr A 1454:32–41; doi:10.1016/j.chroma.2016.05.067. PMID: 27264744.

Evans SF, Kobrosly RW, Barrett ES, Thurston SW, Calafat AM, Weiss B, et al. 2014. Prenatal bisphenol A exposure and maternally reported behavior in boys and girls. Neurotoxicology 45:91–99; doi:10.1016/j.neuro.2014.10.003. PMID: 25307304.

Goñi F, López R, Etxeandia A, Millán E, Amiano P. 2007. High throughput method for the determination of organochlorine pesticides and polychlorinated biphenyls in human serum. J Chromatogr B Analyt Technol Biomed Life Sci 852:15–21; doi:10.1016/j.jchromb.2006.12.049. PMID: 17267299.

Grohs MN, Reynolds JE, Liu J, Martin JW, Pollock T, Lebel C, et al. 2019. Prenatal maternal and childhood bisphenol a exposure and brain structure and behavior of young children. Environ Health 18; doi:10.1186/s12940-019-0528-9. PMID: 31615514.

Harley KG, Gunier RB, Kogut K, Johnson C, Bradman A, Calafat AM, et al. 2013. Prenatal and early childhood bisphenol A concentrations and behavior in school-aged children. Environ Res 126:43–50; doi:10.1016/j.envres.2013.06.004. PMID: 23870093.

Haug LS, Sakhi AK, Cequier E, Casas M, Maitre L, Basagana X, et al. 2018. In-utero and childhood chemical exposome in six European mother-child cohorts. Environ Int 121:751–763; doi:10.1016/j.envint.2018.09.056. PMID: 30326459.

Haug LS, Thomsen C, Becher G. 2009. A sensitive method for determination of a broad range of perfluorinated compounds in serum suitable for large-scale human biomonitoring. J Chromatogr A 1216:385–393; doi:10.1016/j.chroma.2008.10.113. PMID: 19026423.

Koponen J, Rantakokko P, Airaksinen R, Kiviranta H. 2013. Determination of selected perfluorinated alkyl acids and persistent organic pollutants from a small volume human serum sample relevant for epidemiological studies. J Chromatogr A 1309:48–55; doi:10.1016/j.chroma.2013.07.064. PMID: 23972455.

Li F, Yang F, Li D-K, Tian Y, Miao M, Zhang Y, et al. 2020. Prenatal bisphenol A exposure, fetal thyroid hormones and neurobehavioral development in children at 2 and 4 years: A prospective cohort study. Sci Total Environ 722:137887; doi:10.1016/j.scitotenv.2020.137887. PMID: 32197165.

Maitre L, de Bont J, Casas M, Robinson O, Aasvang GM, Agier L, et al. 2018. Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. BMJ Open 8:e021311; doi:10.1136/bmjopen-2017-021311. PMID: 30206078.

Manzano-Salgado CB, Casas M, Lopez-Espinosa M-J, Ballester F, Basterrechea M, Grimalt JO, et al. 2015. Transfer of perfluoroalkyl substances from mother to fetus in a Spanish birth cohort. Environ Res 142:471–478; doi:10.1016/j.envres.2015.07.020. PMID: 26257032.

Perera F, Nolte ELR, Wang Y, Margolis AE, Calafat AM, Wang S, et al. 2016. Bisphenol A exposure and symptoms of anxiety and depression among inner city children at 10-12 years of age. Environ Res 151:195–202; doi:10.1016/j.envres.2016.07.028. PMID: 27497082.

Perera F, Vishnevetsky J, Herbstman JB, Calafat AM, Xiong W, Rauh V, et al. 2012. Prenatal bisphenol A exposure and child behavior in an inner-city cohort. Environ Health Perspect 120:1190–1194; doi:10.1289/ehp.1104492. PMID: 22543054.

Philippat C, Mortamais M, Chevrier C, Petit C, Calafat AM, Ye X, et al. 2012. Exposure to phthalates and phenols during pregnancy and offspring size at birth. Environ Health Perspect 120:464–470; doi:10.1289/ehp.1103634. PMID: 21900077.

Philippat C, Nakiwala D, Calafat AM, Botton J, De Agostini M, Heude B, et al. 2017. Prenatal exposure to nonpersistent endocrine disruptors and behavior in boys at 3 and 5 years. Environ Health Perspect 125:097014; doi:10.1289/EHP1314. PMID: 28937960.

Poothong S, Lundanes E, Thomsen C, Haug LS. 2017. High throughput online solid phase extraction-ultra high performance liquid chromatography-tandem mass spectrometry method for polyfluoroalkyl phosphate esters, perfluoroalkyl phosphonates, and other perfluoroalkyl substances in human serum, plasma, and whole blood. Anal Chim Acta 957:10–19; doi:10.1016/j.aca.2016.12.043. PMID: 28107829.

Ramon R, Murcia M, Aguinagalde X, Amurrio A, Llop S, Ibarluzea J, et al. 2011. Prenatal mercury exposure in a multicenter cohort study in Spain. Environ Int 37:597–604; doi:10.1016/j.envint.2010.12.004. PMID: 21239061.

Rodushkin I, Ödman F, Olofsson R, Axelsson MD. 2000. Determination of 60 elements in whole blood by sector field inductively coupled plasma mass spectrometry. J Anal At Spectrom 15:937–944; doi:10.1039/B003561K.

Roen EL, Wang Y, Calafat AM, Wang S, Margolis A, Herbstman J, et al. 2015. Bisphenol A exposure and behavioral problems among inner city children at 7–9 years of age. Environ Res 142:739–745; doi:10.1016/j.envres.2015.01.014. PMID: 25724466.

Sabaredzovic A, Sakhi AK, Brantsæter AL, Thomsen C. 2015. Determination of 12 urinary phthalate metabolites in Norwegian pregnant women by core-shell high performance liquid chromatography with on-line solid-phase extraction, column switching and tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 1002:343–352; doi:10.1016/j.jchromb.2015.08.040. PMID: 26355271.

Sakhi AK, Sabaredzovic A, Papadopoulou E, Cequier E, Thomsen C. 2018. Levels, variability and determinants of environmental phenols in pairs of Norwegian mothers and children. Environ Int 114:242–251; doi:10.1016/j.envint.2018.02.037. PMID: 29524920.

Tamayo-Uria I, Maitre L, Thomsen C, Nieuwenhuijsen MJ, Chatzi L, Siroux V, et al. 2019. The early-life exposome: description and patterns in six European countries. Environ Int 123:189–200; doi:10.1016/j.envint.2018.11.067. PMID: 30530161.

Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ. 2012. Heavy metals toxicity and the environment. EXS 101:133–164; doi:10.1007/978-3-7643-8340-4_6. PMID: 22945569.

Valvi D, Monfort N, Ventura R, Casas M, Casas L, Sunyer J, et al. 2015. Variability and predictors of urinary phthalate metabolites in Spanish pregnant women. Int J Hyg Environ Health 218:220–231; doi:10.1016/j.ijheh.2014.11.003. PMID: 25558797.

Zerbo O, Iosif A-M, Delwiche L, Walker C, Hertz-Picciotto I. 2011. Month of conception and risk of autism. Epidemiology 22:469–475; doi:10.1097/EDE.0b013e31821d0b53. PMID: 21543984.