Environmental Science & Technology

Article

pubs.acs.org/est

The Pregnancy Exposome: Multiple Environmental Exposures in the INMA-Sabadell Birth Cohort

Oliver Robinson,*[,†,‡,§] Xavier Basagaña,[†,‡,§] Lydiane Agier,[∥] Montserrat de Castro,[†,‡,§] Carles Hernandez-Ferrer,[†,‡,§] Juan R. Gonzalez,[†,‡,§] Joan O. Grimalt,[⊥] Mark Nieuwenhuijsen,[†,‡,§] Jordi Sunyer,[†,‡,§] Rémy Slama,[∥] and Martine Vrijheid[†,‡,§]

[†]Centre for Research in Environmental Epidemiology (CREAL), Doctor Aiguader, 88, E-08003 Barcelona, Spain

[‡]Universitat Pompeu Fabra (UPF), Plaça de la Mercè, 10-12, 08002, Barcelona, Spain

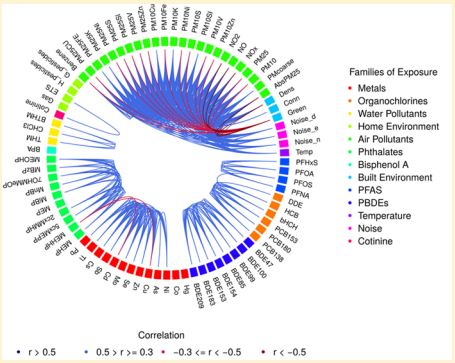
[§]CIBER Epidemiología y Salud Pública (CIBERESP), Casanova, 143, 5 Planta, 08036, Barcelona, Spain

[∥]Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, Inserm and Univ. Grenoble-Alpes, U823 (IAB) Joint Research Center, Grenoble, France

[⊥]Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research, Jordi Girona 18-26, 08034, Barcelona, Spain

Supporting Information

ABSTRACT: The "exposome" is defined as "the totality of human environmental exposures from conception onward, complementing the genome" and its holistic approach may advance understanding of disease etiology. We aimed to describe the correlation structure of the exposome during pregnancy to better understand the relationships between and within families of exposure and to develop analytical tools appropriate to exposome data. Estimates on 81 environmental exposures of current health concern were obtained for 728 women enrolled in The INMA (INfancia y Medio Ambiente) birth cohort, in Sabadell, Spain, using biomonitoring, geospatial modeling, remote sensors, and questionnaires. Pair-wise Pearson's and polychoric correlations were calculated and principal components were derived. The median absolute correlation across all exposures was 0.06 (5th−95th centiles, 0.01−0.54). There were strong levels of correlation within families of exposure (median = 0.45, 5th−95th centiles, 0.07−0.85). Nine exposures (11%) had a correlation higher than 0.5 with at least one exposure outside their exposure family. Effectively all the variance in the data set (99.5%) was explained by 40 principal components. Future exposome studies should interpret exposure effects in light of their correlations to other exposures. The weak to moderate correlation observed between exposure families will permit adjustment for confounding in future exposome studies.



INTRODUCTION

Environmental chemical and physical exposures during fetal or early life have been associated with adverse fetal growth and with developmental neurotoxic, immunotoxic, and obesogenic effects in children, although for many of these associations evidence has been classified as limited or inadequate.[1−4] These are highly complex chronic pathologies, and it is hypothesized that improved understanding of how environmental risk factors coexist and interact during early life can help elucidate their causes.[5−7] It is clear that, up to now, the environment and child health field has almost uniquely focused on single exposure-health effect relationships; there is no global view of how various types of exposures coexist and jointly impact health. The concept of the "exposome"[8] has attracted growing interest in recent years and is defined as the totality of human environmental (i.e., nongenetic) exposures from conception onward, complementing the genome. It is hoped that through the use of holistic and data-driven approaches pioneered in the genomics fields, similar advances can be made in understanding the environmental component of disease etiology. Since the developing fetus is particularly vulnerable to potential environmental hazards and since exposures during the critical *in utero* period may have a lifetime impact, the pregnancy period is an important starting point in characterizing the life course exposome.[9]

Implementing the exposome concept however poses a number of challenges. First, full measurement of the exposome at even a single time point is probably impossible. While "top-down" measurement of exposome signals, either through measurement of the global internal biological response using molecular 'omic technologies or the untargeted analysis of chemicals present in biological samples, addresses this issue to some extent, current analytical technology is not sensitive or

Received: April 8, 2015
Revised: July 9, 2015
Accepted: July 13, 2015
Published: July 13, 2015

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

Table 1. Exposure Families Included in Analysis

exposure family	no. of exposures measured	no. of women	matrix	sampling time	method	ref
PFAS	4	433	serum	first trimester	HPLC-MS	25
organochlorines	6	637			GC-MS	23
PBDEs	8	242	colostrum	birth	GC-MS	39
metals	13	range: 243–489	urine and cord blood (for mercury)	average of first and third trimester spot urines and birth (for mercury)	Q-ICP-MS, AAS (for mercury)	26, 28
phthalates	10	391	urine	average of first and third trimester spot urines	HPLC-MS	29
bisphenol A	1	497			HPLC-MS	29
water pollutants	3	561	home address and questionnaire	pregnancy average	model and interview	31
home environment	4	616	questionnaire	third trimester	interview	21, 32, 33
cotinine	1	597	urine		LC-MS	32
air pollutants	24	range: 573–611			land use regression models	19, 20, 22
built environment	3	range: 477–720	home address	pregnancy average	calculation from maps. Landsat imagery (normalized vegetation difference Index).	34, 36, 37
noise	3	631			municipal strategic noise maps	35
surface temperature	1	728			Landsat imagery (thermal band)	38

flexible enough to capture and identify the components of the exposome in a single analytical sweep.[10,11] Therefore, a complementary approach is to construct the exposome from the "bottom-up" using existing tools of exposure assessment such as biomonitoring across various analytical platforms, geospatial modeling, and questionnaires. Although this is a laborious process it should be recognized that even partial exposome coverage will be valuable and other 'omic-wide scans such as the genome wide association studies (GWAS) rely on incomplete coverage, supplemented with imputation based on resources such as haplotype databases.[12,13] Second, appropriate statistical tools, which provide sufficient sensitivity and specificity in the face of the high dimensionality and dense correlations inherent in exposome data, will be required for exposome-health association studies. Understanding what a typical exposome looks like, including the structure of correlations between and within groups of exposure is an important first step in planning the optimal use of both targeted exposome measurements and statistical analyses. Some of the first exposome studies[1415] have used the cross-sectional National Health and Nutrition Examination Survey (NHANES) biomonitoring data, which has provided reference values of contamination levels to a wide range of both environmental pollutants and nutrients.[16] Patel and Ioannidis[17] recently described the correlation structure of the NHANES data set, giving ranges of absolute correlation for each exposure group analyzed, providing important information on part of the exposome structure of the general U.S. population. Furthermore, they proposed that transparent knowledge of the correlation structure of a data set is required to best interpret reported results using that data set, which over the course of particular study may number hundreds of publications.

INMA (INfancia y Medio Ambiente) is a birth cohort study in seven regions of Spain that aims to examine the role of environmental pollutants during pregnancy and early childhood in relation to child growth and development.[18] The INMA Sabadell subcohort, situated in Catalonia, has already described levels of exposure to a range of environmental factors during pregnancy including outdoor and indoor air pollution,[19−22] persistent organic pollutants (POPs,[23] brominated flame retardants,[24] perfluoroalkyl substances (PFAS,[25]), metals,[26−28] phenols and phthalates,[29,30] disinfection byproducts in water,[31] environmental tobacco smoke,[32] insecticides,[33] and green spaces.[34] Here we present an analysis of relationships within and between important groups of environmental exposure among pregnant Spanish women with the aim of better understanding the correlation structure of an important part of the "pregnancy exposome".

METHODS

A full description of the project protocol has been previously described.[18] Briefly, during 2004−2006, pregnant women (N = 728) from the general population were recruited at the first trimester routine antenatal care visit in the main public hospital or health center of reference, using the following inclusion criteria: Women had to be at least 16 years old, intend to deliver in the reference hospital, have a singleton pregnancy with no assisted conception, and have no problems with communication. The study was conducted with the approval of the hospital ethics committee, and written informed consent was obtained from all women.

Estimates on 81 exposures covering the pregnancy period were collated into a single data set. Exposures were selected on the basis of availability from current or ongoing INMA studies. Biomonitoring data included organochlorines (including pesticides and polychlorinated biphenyls (PCBs)) and PFAS in serum; mercury in cord blood; polybrominated diphenyl ethers (PBDEs) in breast milk; metals, phthalates, bisphenol A in urine (averaged over measurements in samples collected during the first and third trimesters) ; and cotinine in urine (from third trimester) (Table 1). Biomarker measurements where the analyte was nondetectable in over 85% of samples (including lead in cord blood, PCB congeners 28, 52, 101, and 118, dichlorodiphenyltrichloroethane (DDT), and PBDE congeners 17, 28, 71, 66, 138, and 190) were excluded from the analysis. Geospatial modeling and remote sensing data included air pollutants (including nitrogen oxides, particulate

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

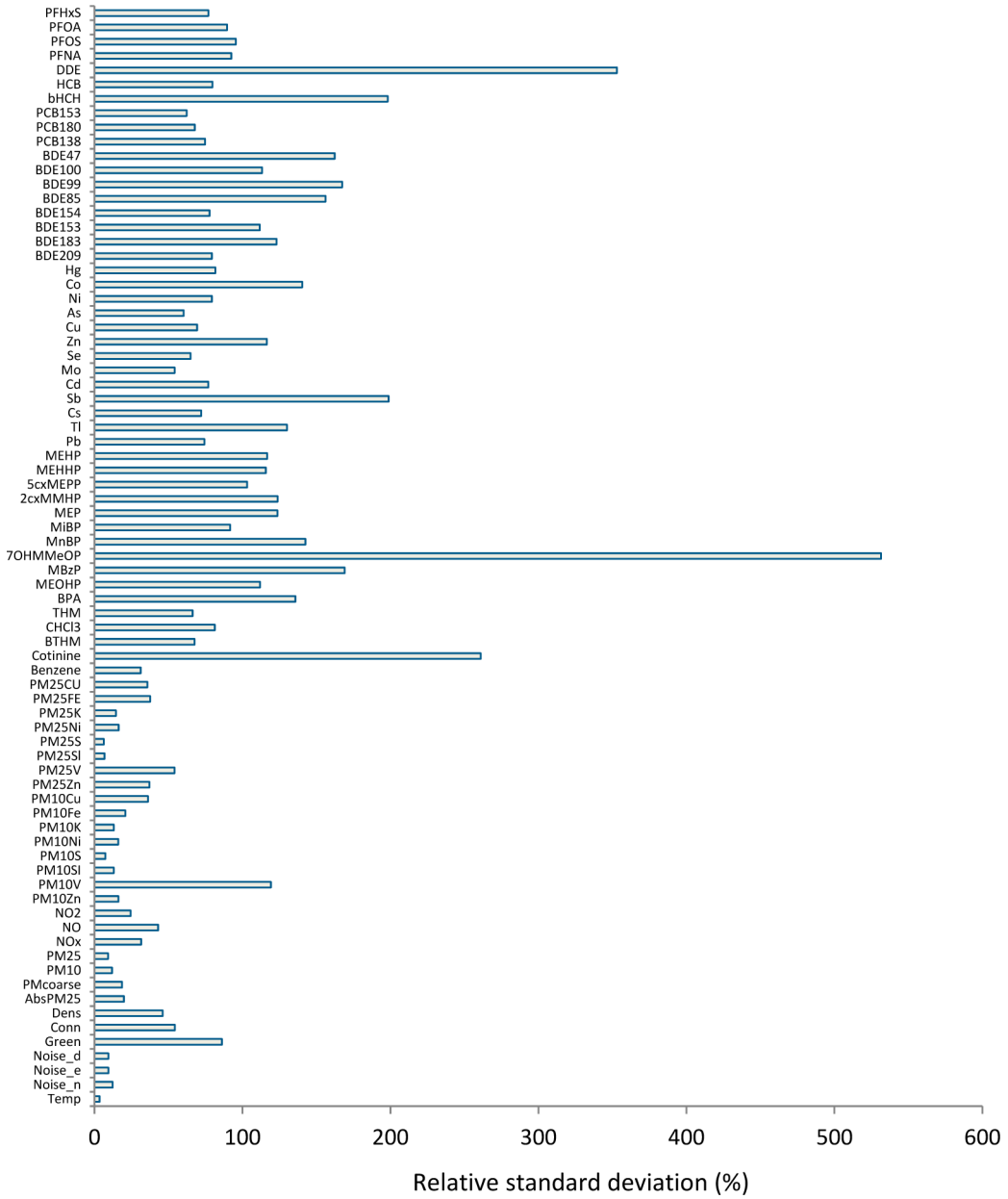


Figure 1. Relative standard deviation (standard deviation/mean) for each continuous exposure. Abbreviations for all exposures are shown in Supplementary Table S1 in the Supporting Information.

matter (of diameter less than 2.5 μm ($PM_{2.5}$), less than 10 μm (PM_{10}), and between 2.5 and 10 μm (PM_{coarse})) $PM_{2.5}$ absorbance (a measure of black carbon) and various elemental fractions of $PM_{2.5}$ and PM_{10}), the built environment (building density, street connectivity, and green spaces), noise (averages over day, evening, and night), and land surface temperature. Questionnaire data, collected by trained interviewers during the third trimester, included four home environment related binary variables including gas cooking, home and garden pesticide use, and environmental tobacco smoke exposure. Water use habits collected by questionnaire were combined with modeled levels of disinfection byproducts (total trihalomethanes, brominated trihalomethanes, and chloroform) in the residential water supply to calculate daily ingestion. References for the original studies and methods used are shown in Table 1. Additionally in this study, estimates on noise, surface temperate, building density, and street connectivity were assigned to the home address of participants within the ArcGIS platform (ESRI ArcMap TM 10.0, ArcGIS Desktop 10 Service Pack 4, spatialite v.4.11). Noise exposures were obtained from the strategic noise maps for Sabadell produced by the Generalitat de Catalunya under the European Noise Directive.[35] Total building area and number of street intersections within a 100 m radius buffer were calculated from topographical[36] and road network maps.[37] The radiometric surface temperature was calculated from LANDSAT thermal imagery within a 50 m radius buffer from the home address.[38] Exposures were grouped into families depending on their structure (for individually measured biomarkers) or source (other exposures) (Table 1).

Continuous variables were log-transformed to give a normal distribution. Biomarker measurements below the detection limit were imputed using distribution-based multiple imputation.[40] The proportion of biomarker measurements below the detection limit are shown in Supplementary Table S1

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

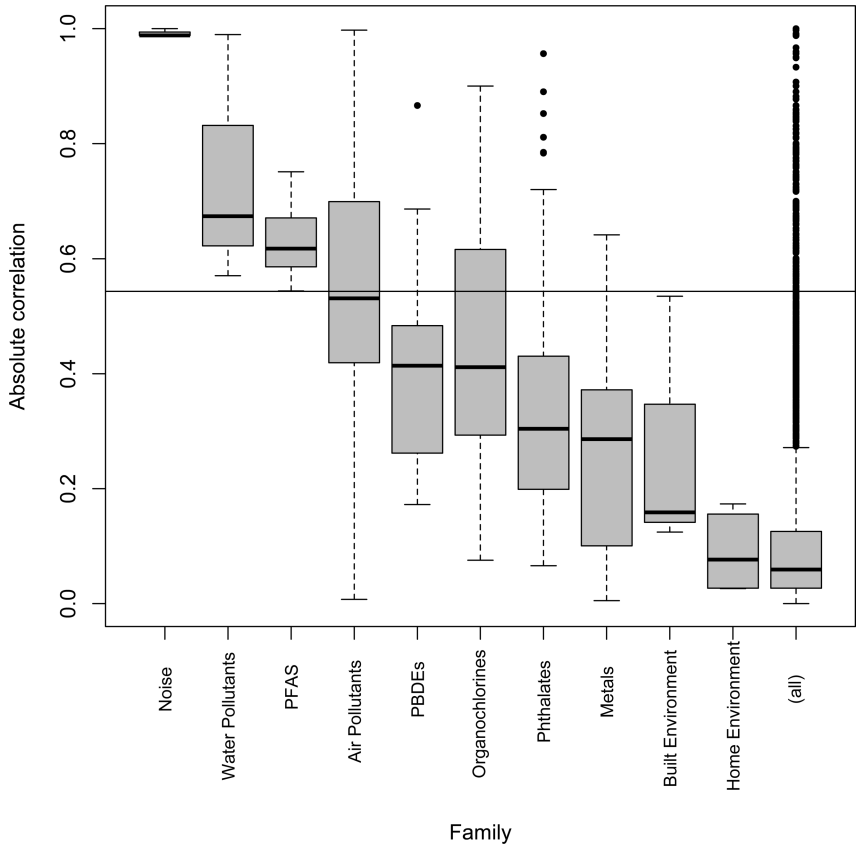


Figure 2. Pairwise correlations (absolute value) within families of exposure (for families with more than one exposure). Boxes illustrate interquartile range (IQR) with median displayed as a thick horizontal black line in the middle of the box. The whiskers extend to the most extreme data point, which is no more than 1.5 times the IQR from the box. Outliers are shown in actual points. "(all)" denotes correlation across all pairs of variables available; the horizontal line on the graph denotes the 95th percentile of these absolute correlations.

(Supporting Information). Pair-wise Pearson's correlations (for continuous variables) and polychoric correlations (for correlations involving binary variables) between each individual exposures were calculated to produce a correlation matrix. Heat map and circos plots were made to display the correlations. Principal components were then derived directly from the correlations. All analyses were conducted in the R software environment (http://www.r-project.org/index.html).

RESULTS

The number of women with available exposure estimates ranged from 242 for the PBDEs to 728 women for temperature (mean number of women per exposure, 501). All exposures, along with summary statistics of their levels, are listed in the Supplementary Table S1 in the Supporting Information. The percentage relative standard deviation (standard deviation/mean) for each exposure ranged from 3% for surface temperature at the home address to 531% for mono (4-methyl-7-hydroxyoctyl) phthalate (7OHMMeOP) (Figure 1, Supplementary Table S1 in the Supporting Information), with a mean relative standard deviation across all exposures of 84%. The mean correlation (r) across all exposures was 0.08, with a standard deviation of 0.21 (median =0.02; 5−95th centiles = −0.12 to 0.54). The mean *absolute* correlation was 0.13, with standard deviation of 0.18 and range 0.00 to 1.00 (median = 0.06; 5−95th centiles = 0.01−0.54).

There were strong levels of correlation within families of exposure with absolute correlations strongest among the noise indicators (median r = 0.99) and weakest among the home environment exposures (median r = 0.08) (Figure 2). The water disinfection byproducts and air pollutants had strong median levels of absolute correlation (r = 0.67 and 0.53, respectively) although with large ranges. The four PFOA compounds had the strongest median absolute correlation (r = 0.62) of the individually measured biomarkers. The other biomarker families, PDBEs, phthalates, metals, and organochlorines, all had median absolute correlations below 0.5, reflecting their more diverse sources. However, some pairwise correlations within each of these families were above 0.5. The built environment measures showed lower levels of correlations between them, with an absolute median correlation of 0.16. The strongest correlation within the home environment exposures was between use of home and garden pesticides (r = 0.16). Overall, the median of all within-family absolute correlations was 0.45 (5th−95th centiles, 0.07−0.85).

The correlation heatmap (Figure 3) displays the linkage across all exposures by their correlation. "Blocks" of high correlation within families of exposure were observed along the main diagonal of the heat map, with certain groups such as the organochlorines and phthalate metabolites showing less dense within-family correlations than more closely linked exposures such as the PFAS. With respect to between family correlations, no exposure had an absolute correlation higher than 0.6 with an exposure outside its family. Nine exposures (11% of all 81 exposures) had an absolute correlation higher than 0.5 with at least one exposure outside its family. These included nighttime noise, which had a correlation of 0.52 with the air concentration of the copper fraction of $PM_{2.5}$; proximity to green spaces,

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

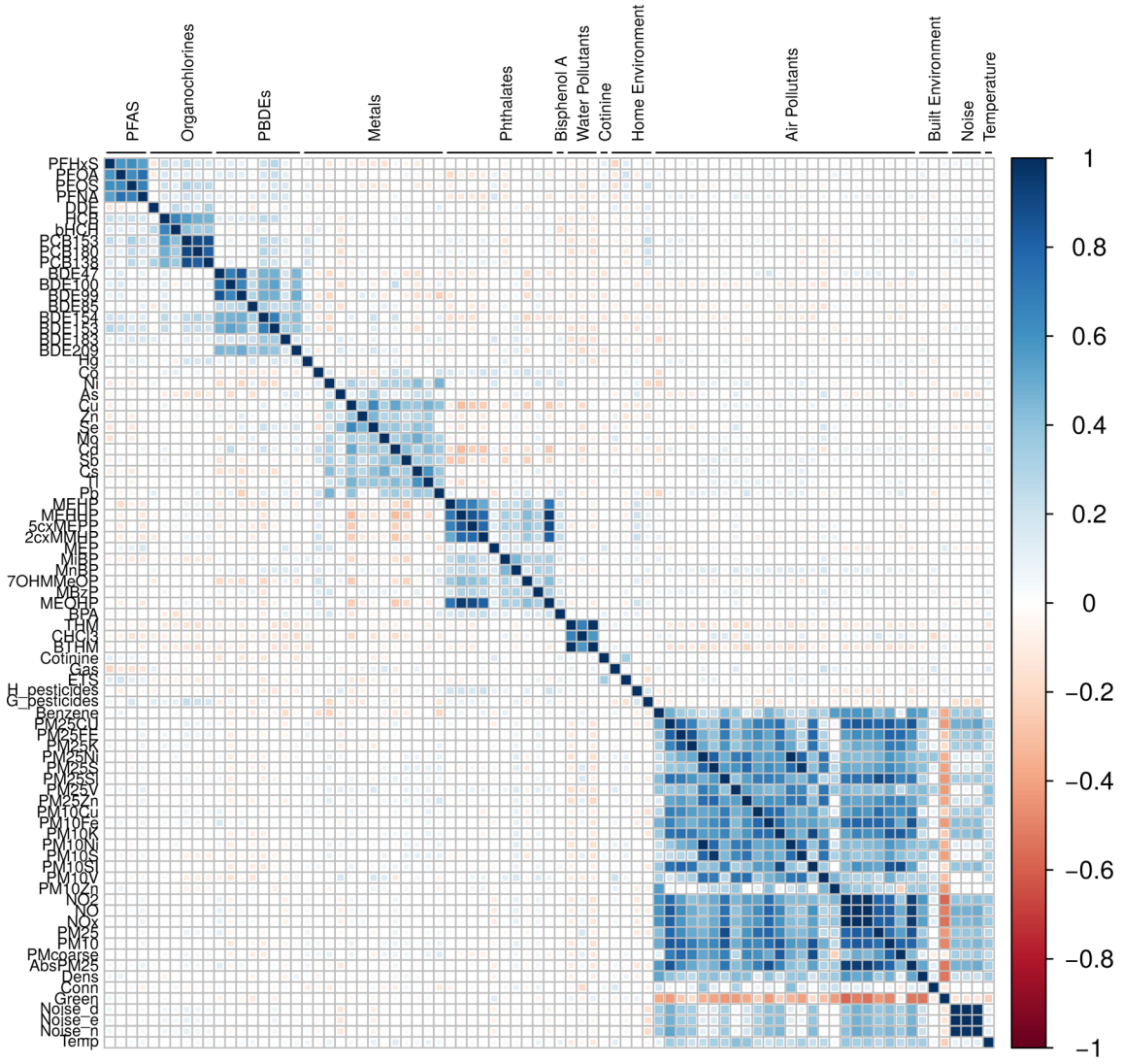


Figure 3. Correlation heatmap, showing pair correlations across all exposures, with blue color indicating positive correlations and red color indicating negative correlations. Abbreviations for all exposures are shown in Supplementary Table S1 in the Supporting Information.

which had negative correlations with nitrogen oxides and $PM_{2.5}$ absorbance; and building density which was positively correlated with benzene, nitrogen oxides, and $PM_{2.5}$ absorbance. A total of 26 exposures (32%) had an absolute correlation higher than 0.4 with at least one exposure outside its family: In addition to further associations between noise variables, green space proximity, building density, and air pollutants, we observed a positive correlation of 0.43 between street connectivity and the nickel fraction of PM_{10}. Street connectivity also had a correlation of 0.39 with the nickel fraction of $PM_{2.5}$ and a correlation of 0.32 with the vanadium fraction of $PM_{2.5}$. Surface temperature had a correlation of 0.41 with the vanadium fraction of $PM_{2.5}$ and correlations above 0.3 with 11 other air pollutants. As may be expected, urinary cotinine was correlated (r = 0.35) with self-reported environmental tobacco smoke exposure. In general, for those exposures measured individually, through biomarker or questionnaire, there was low correlation between exposures in separate families: Only three pairwise correlations between biomarker-measured exposures in separate families were above 0.3 with the strongest correlation observed between perfluorooctanesulfonic acid and PCB-153 (r = 0.32). Overall, the median of all between-family absolute correlations was 0.05 (5th−95th centiles, 0.01−0.23).

Only three principal components were required to explain 50% of variance across the whole data set, while six components explained 70% of variance and 22 components explained 95% of the variance (Supplementary Figure S1 in the Supporting Information, Table 2). The components were not solely loaded onto single exposure families. The exposures most strongly loading onto the first component (absolute loading >0.10) were primarily outdoor environment exposures, including all the air pollutants, building density, noise, surface temperature (range of loadings, −0.11 to −0.19), and green spaces (0.16). The second component was composed primarily of positive loadings to the PFAS (0.11−0.15), PBDEs (0.12 to −0.22), hexachlorobenzene (0.14), PCB congeners 153 and 180 (both 0.11), and some metals (0.11−0.17) and negative loadings to the phthalates (monoethyl phthalate (MEP) = −0.08 and others −0.13 to −0.24) and BPA (−0.15), with further contributions from cobalt (−0.11) and home pesticides (−0.17). The third component was composed of positive loadings to all the metals except mercury and cobalt (0.18−0.27) and strong negative loadings to the PFAS, the

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

Table 2. Principal Component (PC) Analysis Showing the Number of Components Required to Explain Percentages of Cumulative Variance by Each Exposure Group and Across All Exposures

		no. of PCs required to explain % of cumulative variance:			
exposure group	no. of variables	50	70	95	99.5
PFAS	4	2	2	3	3
organochlorines	6	1	2	3	4
PBDEs	8	2	3	5	6
metals	13	2	4	9	11
phthalates	10	1	3	5	7
bisphenol A	1				
water pollutants	3	1	1	2	2
cotinine	1				
home environment	4	2	2	3	3
air pollutants	24	2	2	5	10
built environment	3	1	2	2	2
noise	3	2	1	1	1
temperature	1				
all	81	3	6	22	40

organochlorines (except DDE), five of the PBDEs, and three of the phthalates (−0.10 to −0.18). A total of 99.5% of variance in the data set, which may be considered effectively all variance, was explained by 40 components. Within each exposure family, only one component was needed to explain 99.5% of variance among the three noise variables, 10 components were needed to explain 99.5% of variance among the 24 air pollution variables while 11 components were needed to explain 99.5% of variance among the 13 metals (Table 2). Principal component loadings can be found in Supplementary Table S2 in the Supporting Information.

DISCUSSION

Since its initiation in 2004, the INMA Sabadell birth cohort has measured exposure to many of the most important environmental factors of current concern to child health, providing a wide range of exposure estimates covering the *in utero* period for a substantial number of Spanish women. This has provided a rich resource of environmental data for the study of longitudinal health outcomes in children and has now allowed a first picture of the structure of an important piece of the pregnancy exposome, a key starting point in constructing a life course exposome.

The presented correlation structure enables improved interpretation of results reported by both the INMA Sabadell birth cohort and in epidemiological studies in general. As with other reported exposure correlations[17,41] we find strong levels of correlations within families of exposure (grouped by structure or source) and therefore results reported for single exposures need to be interpreted in light of their correlations to other exposures within their respective families. An increasing number of studies are now including multiple within family exposures[9] and it has long been recognized that air pollution studies should consider multiple pollutants simultaneously, although in the presence of high correlation it becomes difficult to disentangle the effects of each pollutant.[42] We see weak levels of correlation between the families of chemical exposures measured in individual women and with other families of exposure. This provides confidence that reported results for biomarker exposure estimates are not confounded by correlation with other unreported exposures and provides scope for epidemiological studies to separate the effects of each exposure group. We do however see stronger levels of correlation between families of exposures encountered in the outdoor environment such as air pollutants, noise, temperature, and the built environment indicating that studies focusing on one of these families should be interpreted with caution. However, the range of between-family correlations found in this study, all lower than 0.6, would allow disentangling of their effects if all exposures have been measured. Future studies should consider these families of outdoor exposures in combination in order to provide appropriate risk estimates, an approach adopted now by a growing number of studies.[43,44]

In the environment-wide association study (EWAS) approach to exposome analysis, analogous to GWAS, adopted by Patel and colleagues[14] (and they argue, adopted implicitly by research projects with repeated publications), multiple analyses of single pollutants must be adjusted to guard against generation of false positives by a Bonferroni correction or similar. In the presence of correlations between exposures (or linkage disequilibrium between single nucleotide polymorphisms (SNPs) in GWAS), Bonferroni correction would be overly conservative and instead correction should be made based on effective, rather than actual variables. Patel and Ioannidis[17] report that the 530 exposure variables available in the NHANES data set could be reduced to 476 "effective variables" based on the within family correlations following the method of Nyolt.[45] Here for simplicity we have presented the number of principal components required to explain the variances observed for each exposure group and across exposure groups for the whole data set. Following the method of Gao et al.,[46] which was demonstrated to provide more efficient multiple testing correction when there is high linkage disequilibrium (or correlation in the exposome context), the number of effective variables is equivalent to the number of principal components required to explain 99.5% of the variance. Therefore, of the 81 variables analyzed here, we find that there are 40 "effective variables" that explain practically all the variance contained in the data set. A hypothetical EWAS analysis of the INMA Sabadell pregnancy exposome data set using a Bonferroni-type correction may therefore choose a p value threshold of 0.001 (i.e., 0.05/40).

Although the EWAS approach is flexible, other methodologies (reviewed in ref 47) may prove more appropriate. The presented results may be used as a foundation in simulation studies to assess the performance of different statistical models for the analysis of exposome data. However, one difficulty when analyzing associations with health outcomes using this data set, particularly when applying multivariate methods, is missing values. To maintain a breadth of exposures that approaches an "exposome" data set, while also retaining a sufficient number of observations would not be possible, since not all exposures, particularly those derived from biomarkers, were available for all women. A common solution in these situations is to use an imputed data set[41,48] which is a justifiable approach for analyses on large populations, providing certain assumptions hold.[49] Imputation also provides a more general solution to providing the wide coverage of the external exposome required for agnostic exposome wide scans. Since we find that effectively all the variance in this data set could be explained by much fewer principal components than the actual number of variables measured, improved knowledge of the correlation structure of

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

the exposome may allow external exposome assessment based on fewer measured exposures. To return to the GWAS analogy, wide and more cost-effective genome coverage is provided by genotyping of a few hundred thousand SNPs, which provide information on several million base pairs based on imputation. "Hits" detected in imputed regions may then be followed up with deep sequencing to confirm results. The required correlation or haplotype information is provided by consortia who have conducted full genome sequencing on only a relatively small number of individuals representing a range of different ethnicities.[12,13] Although we have presented PCA to describe the underlying dimensions of the data, it may not be the optimal strategy to select the exposures best able to provide a wide coverage of the exposome. Several other variable reduction techniques exist,[50] and the final set of selected variables may vary according to each technique.

Whether this approach is applicable to exposome research will depend on how reproducible the exposome correlation structure is across spatially and temporally distributed populations. Only some of the correlations reported here may be compared to correlations reported in other data sets such as NHANES[17] and the study of Lenters et al.[41] that measured multiple biomarkers among males of reproductive age living in Greenland, Ukraine, and Poland. For instance of the 13 phthalate metabolites measured in NHANES, the median absolute correlation is 0.25, of the four phthalate metabolites reported by Lenters et al. the median absolute correlation is 0.38, and the median absolute correlation of the 10 phthalate metabolites measured here is 0.30. The correlation levels will, however, be very dependent on the analytes chosen within particular families; it is of interest that the median absolute correlation of four PFAS that were measured both in this study and in the study of Lenters et al. was very similar (0.62 and 0.68, respectively). Between-family correlations were generally weak in both the study of Lenters et al. and as reported here. However, Lenters et al. found relatively strong correlations between mercury, the PFAS, and PCB-153, which were absent among the INMA Sabadell women. This is likely explained by the higher levels found for all these chemicals among the studied Greenland population since correlations were not reported separately for the three included regions. Thus, while some aspects of the correlation structure of exposomes of populations around the world are similar, other aspects will depend on the particular environment and lifestyles of the population studied, such as the consumption of large marine animals in the Greenland population. A future "Human Exposome Project" will need to consider measurements in a range of populations with standardized exposure measurements (with respect to analyte, method, and matrix analyzed).

The INMA Sabadell pregnancy exposome data set provides broad coverage of the exposome since indicators of most environmental exposure groups of key current concern are included, covering biomarker measurements of both persistent and nonpersistent pollutants, questionnaire information on personal commercial product use, and geospatial modeled estimates on air and water contaminants. Exposures derived from geospatial models are not included in the NHANES data set but should be included in characterization of the external exposome since they provide information for many exposures for which specific biomarkers are not available. However, we observed overall differences in the variability between those exposures measured through biomarkers and those exposures that were assigned based on address. This may be problematic for exposome analyses since statistical models may have reduced sensitivity to detect health effects from those exposures with lower variability. Biomarker measurements may have higher between subject variability because they incorporate information regarding both prevalence in the environment and personal behavior. The precision of the geospatially derived estimates will be improved when supplemented with information about how individuals move through their environment, now becoming available from smartphones.[51] Similarly the binary estimates on personal product use would be improved with the use of more detailed questionnaires. Further limitations to the current analysis include preselection of the included analytes which may limit their utility in truly agnostic exposome analyses.[11] Future measurements of the external exposome may consider choosing indicator exposures to provide the widest exposome coverage (i.e., representative of most exposure groups) rather than those of most regulatory concern. Furthermore, one must consider that parts of the correlation structure presented here are composed of analytical variability; those exposures measured using the same analytical platform may show greater within platform correlation compared to those measured on other platforms, obscuring "true" biological variability and correlation. Outdoor exposure models constructed from the same variables, such as traffic density, may similarly show inflated correlations. A final important limitation is the different degrees of exposure misclassification between exposures. As with all exposure assessment, efforts are needed to improve assessment of each exposure. Misclassification may be high for nonpersistent exposures such as BPA since it is known that within person variability for these compounds is high.[52] Despite addressing this to some extent using the average of urinary measurements at two time points, exposure misclassification will be greater than for other exposures such as air pollution for which routine monitoring can provide daily and relatively accurate exposure estimates at the address level.[20] Combined analyses of exposures with differential measurement error may decrease the accuracy of joint effect estimates, with the effect of the well-measured exposures dominating the effect estimates for correlated but less well-measured exposures.[53]

These limitations may only be overcome with the development of an "exposome chip" for a single exposome analysis or similarly, the concurrent analysis of the "top-down" exposome and its relationship to the "bottom-up" exposome presented here.[11,54] Potential future analyses in the INMA Sabadell exposome data set involve the inclusion of other parts of the external exposome such as diet, physical activity, and drug use and more general social and economic factors. This more complete external exposome could also be examined in relation to available measures of the internal exposome (i.e., the biological response and endogenously derived exposures) that include metabolome, DNA methylation, and inflammatory markers. Assessment of the internal exposome using molecular 'omic technologies may allow more appropriate grouping of external exposures based on shared toxicogenomic effects. Furthermore, ongoing European-wide research projects such as the HELIX project[55] will provide broad coverage in a large numbers of subjects of both the external and internal exposomes and test the utility of both the "bottom-up" and the "top-down" approaches.

In summary, the correlation analysis presented here of multiple environmental exposures among pregnant women provides a first picture of the structure of the exposome during

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

the crucial *in utero* period. This information will aid interpretation of reported findings from epidemiological studies in general and inform future analyses of the exposome.

■ ASSOCIATED CONTENT

Supporting Information

Full correlation matrix in Excel format; summary descriptions of exposure levels and acronyms for all exposures; figure of cumulative variance explained by principal component analysis across whole data set; exposure loadings for up to the first 10 components for principal component analysis across whole data set and for each exposure family. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b01782.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: orobinson@creal.cat; phone: +34 932 14 7362.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank all INMA researchers on whose previous work this study is based. We thank all the INMA study participants and the members of the HELIX-EXPOsOMICs statistical working group who provided helpful discussions This research received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 308333—the HELIX project. The INMA Sabadell cohort was funded by grants from Instituto de Salud Carlos III (Red INMA G03/176, CB06/02/0041), Spanish Ministry of Health (FIS-PI041436, FIS- PI081151), Generalitat de Catalunya (CIRIT 1999SGR 00241), Fondo de Investigacion Sanitaria (ISCIII: FIS-PI12/01890), Fundació La Marató De TV3 (090430), and Recercaixa (2010ACUP 00349). We acknowledge the following studies, funded by the European Community's Sixth and Seventh Framework Programmes, for generating exposure estimates used in this analysis: The ESCAPE study (Grant 211250) provided estimates on exposure to air pollutants; the HIWATE study (Contract No. Food-CT-2006-036224) provided estimates on exposure to water disinfection byproducts; and the PHENOTYPE study (Grant 282996) provided estimates on green spaces.

■ REFERENCES

(1) Gascon, M.; Morales, E.; Sunyer, J.; Vrijheid, M. Effects of persistent organic pollutants on the developing respiratory and immune systems: A systematic review. *Environ. Int.* **2013**, *52C*, 51−65.

(2) Wigle, D. T.; Arbuckle, T. E.; Turner, M. C.; Berube, A.; Yang, Q.; Liu, S.; Krewski, D. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. *J. Toxicol. Environ. Health, Part B* **2008**, *11* (5−6), 373−517.

(3) Grandjean, P.; Landrigan, P. J. Neurobehavioural effects of developmental toxicity. *Lancet Neurol.* **2014**, *13* (3), 330−8.

(4) Thayer, K. A.; Heindel, J. J.; Bucher, J. R.; Gallo, M. A. Role of environmental chemicals in diabetes and obesity: a National Toxicology Program workshop review. *Environ. Health Perspect* **2012**, *120* (6), 779−89.

(5) Van den Bergh, B. R. Developmental programming of early brain and behaviour development and mental health: a conceptual framework. *Dev. Med. Child Neurol.* **2011**, *53* (Suppl 4), 19−23.

(6) Trasande, L.; Cronk, C.; Durkin, M.; Weiss, M.; Schoeller, D. A.; Gall, E. A.; Hewitt, J. B.; Carrel, A. L.; Landrigan, P. J.; Gillman, M. W. Environment and obesity in the National Children's Study. *Environ. Health Perspect.* **2009**, *117* (2), 159−66.

(7) Bousquet, J.; Anto, J.; Auffray, C.; Akdis, M.; Cambon-Thomsen, A.; Keil, T.; Haahtela, T.; Lambrecht, B. N.; Postma, D. S.; Sunyer, J.; Valenta, R.; Akdis, C. A.; Annesi-Maesano, I.; Arno, A.; Bachert, C.; Ballester, F.; Basagana, X.; Baumgartner, U.; Bindslev-Jensen, C.; Brunekreef, B.; Carlsen, K. H.; Chatzi, L.; Crameri, R.; Eveno, E.; Forastiere, F.; Garcia-Aymerich, J.; Guerra, S.; Hammad, H.; Heinrich, J.; Hirsch, D.; Jacquemin, B.; Kauffmann, F.; Kerkhof, M.; Kogevinas, M.; Koppelman, G. H.; Kowalski, M. L.; Lau, S.; Lodrup-Carlsen, K. C.; Lopez-Botet, M.; Lotvall, J.; Lupinek, C.; Maier, D.; Makela, M. J.; Martinez, F. D.; Mestres, J.; Momas, I.; Nawijn, M. C.; Neubauer, A.; Oddie, S.; Palkonen, S.; Pin, I.; Pison, C.; Rance, F.; Reitamo, S.; Rial-Sebbag, E.; Salapatas, M.; Siroux, V.; Smagghe, D.; Torrent, M.; Toskala, E.; van Cauwenberge, P.; van Oosterhout, A. J.; Varraso, R.; von Hertzen, L.; Wickman, M.; Wijmenga, C.; Worm, M.; Wright, J.; Zuberbier, T. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy* **2011**, *66* (5), 596−604.

(8) Wild, C. P. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol., Biomarkers Prev.* **2005**, *14* (8), 1847−50.

(9) Robinson, O.; Vrijheid, M. The Pregnancy Exposome. *Current Environmental Health Reports* **2015**, *2* (2), 204−213.

(10) Rappaport, S. M.; Barupal, D. K.; Wishart, D.; Vineis, P.; Scalbert, A. The blood exposome and its role in discovering causes of disease. *Environ. Health Perspect.* **2014**, *122* (8), 769−74.

(11) Lioy, P. J.; Rappaport, S. M. Exposure science and the exposome: an opportunity for coherence in the environmental health sciences. *Environ. Health Perspect.* **2011**, *119* (11), A466−7.

(12) The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **2010**, *467* (7319), 1061−1073.

(13) The International HapMap Consortium. The International HapMap Project. *Nature* **2003**, *426* (6968), 789−796.

(14) Patel, C. J.; Bhattacharya, J.; Butte, A. J. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* **2010**, *5* (5), e10746.

(15) Patel, C. J.; Cullen, M. R.; Ioannidis, J. P.; Butte, A. J. Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels. *Int. J. Epidemiol.* **2012**, *41* (3), 828−43.

(16) CDC National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm (accessed December 20, 2014).

(17) Patel, C. J.; Ioannidis, J. P. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J. Epidemiol. Community Health* **2014**, *68* (11), 1096−100.

(18) Guxens, M.; Ballester, F.; Espada, M.; Fernández, M. F.; Grimalt, J. O.; Ibarluzea, J.; Olea, N.; Rebagliato, M.; Tardón, A.; Torrent, M.; Vioque, J.; Vrijheid, M.; Sunyer, J. Cohort Profile: the INMA–INfancia y Medio Ambiente–(Environment and Childhood) Project. *International journal of epidemiology* **2012**, *41*, 930−40.

(19) Aguilera, I.; Sunyer, J.; Fernandez-Patier, R.; Hoek, G.; Aguirre-Alfaro, A.; Meliefste, K.; Bomboi-Mingarro, M. T.; Nieuwenhuijsen, M. J.; Herce-Garraleta, D.; Brunekreef, B. Estimation of outdoor NO(x), NO(2), and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environ. Sci. Technol.* **2008**, *42* (3), 815−21.

(20) Wang, M.; Beelen, R.; Basagana, X.; Becker, T.; Cesaroni, G.; de Hoogh, K.; Dedele, A.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Forastiere, F.; Galassi, C.; Grazuleviciene, R.; Hoffmann, B.; Heinrich, J.; Iakovides, M.; Kunzli, N.; Korek, M.; Lindley, S.; Molter, A.; Mosler, G.; Madsen, C.; Nieuwenhuijsen, M.; Phuleria, H.; Pedeli, X.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Stempfelet, M.; Tsai, M. Y.; Lanki, T.; Udvardy, O.; Varro, M. J.; Wolf, K.;

Weinmayr, G.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Evaluation of land use regression models for NO2 and particulate matter in 20 European study areas: the ESCAPE project. *Environ. Sci. Technol.* **2013**, *47* (9), 4357−64.
(21) Vrijheid, M.; Martinez, D.; Aguilera, I.; Bustamante, M.; Ballester, F.; Estarlich, M.; Fernandez-Somoano, A.; Guxens, M.; Lertxundi, N.; Martinez, M. D.; Tardon, A.; Sunyer, J.; Project, I. Indoor air pollution from gas cooking and infant neurodevelopment. *Epidemiology* **2012**, *23* (1), 23−32.
(22) de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Eeftens, M.; Eriksen, K.; Eriksson, C.; Fischer, P.; Grazuleviciene, R.; Gryparis, A.; Hoffmann, B.; Jerrett, M.; Katsouyanni, K.; Iakovides, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Mosler, G.; Nador, G.; Nieuwenhuijsen, M.; Pershagen, G.; Peters, A.; Phuleria, H.; Probst-Hensch, N.; Raaschou-Nielsen, O.; Quass, U.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Schwarze, P.; Tsai, M. Y.; Yli-Tuomi, T.; Varro, M. J.; Vienneau, D.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Development of land use regression models for particle composition in twenty study areas in Europe. *Environ. Sci. Technol.* **2013**, *47* (11), 5778−86.
(23) Valvi, D.; Mendez, M. A.; Garcia-Esteban, R.; Ballester, F.; Ibarluzea, J.; Goni, F.; Grimalt, J. O.; Llop, S.; Marina, L. S.; Vizcaino, E.; Sunyer, J.; Vrijheid, M. Prenatal exposure to persistent organic pollutants and rapid weight gain and overweight in infancy. *Obesity* **2014**, *22* (2), 488−96.
(24) Gascon, M.; Vrijheid, M.; Martinez, D.; Forns, J.; Grimalt, J. O.; Torrent, M.; Sunyer, J. Effects of pre and postnatal exposure to low levels of polybromodiphenyl ethers on neurodevelopment and thyroid hormone levels at 4 years of age. *Environ. Int.* **2011**, *37* (3), 605−11.
(25) Manzano-Salgado, C. Perfluoroalkyl substances in matched maternal-cord samples in a Spanish birth cohort: Results of a pilot study. In *ISEE-EUROPE 2014: Young Researchers Conference on Environmental Epidemiology*, Barcelona, Spain, October 20−21, 2014; poster.
(26) Ramon, R.; Murcia, M.; Aguinagalde, X.; Amurrio, A.; Llop, S.; Ibarluzea, J.; Lertxundi, A.; Alvarez-Pedrerol, M.; Casas, M.; Vioque, J.; Sunyer, J.; Tardon, A.; Martinez-Arguelles, B.; Ballester, F. Prenatal mercury exposure in a multicenter cohort study in Spain. *Environ. Int.* **2011**, *37* (3), 597−604.
(27) Fort, M.; Grimalt, J. O.; Casas, M.; Sunyer, J. Interdependence between urinary cobalt concentrations and hemoglobin levels in pregnant women. *Environ. Res.* **2015**, *136* (0), 148−154.
(28) Fort, M.; Cosin-Tomas, M.; Grimalt, J. O.; Querol, X.; Casas, M.; Sunyer, J. Assessment of exposure to trace metals in a cohort of pregnant women from an urban center by urine analysis in the first and third trimesters of pregnancy. *Environ. Sci. Pollut. Res.* **2014**, *21* (15), 9234−41.
(29) Gascon, M.; Casas, M.; Morales, E.; Valvi, D.; Ballesteros-Gomez, A.; Luque, N.; Rubio, S.; Monfort, N.; Ventura, R.; Martinez, D.; Sunyer, J.; Vrijheid, M. Prenatal exposure to bisphenol A and phthalates and childhood respiratory tract infections and allergy. *J. Allergy Clin. Immunol.* **2014**, *135* (2), 370−378.e7.
(30) Casas, L.; Fernandez, M. F.; Llop, S.; Guxens, M.; Ballester, F.; Olea, N.; Irurzun, M. B.; Rodriguez, L. S.; Riano, I.; Tardon, A.; Vrijheid, M.; Calafat, A. M.; Sunyer, J.; Project, I. Urinary concentrations of phthalates and phenols in a population of Spanish pregnant women and children. *Environ. Int.* **2011**, *37* (5), 858−866.
(31) Villanueva, C. M.; Gracia-Lavedan, E.; Ibarluzea, J.; Santa Marina, L.; Ballester, F.; Llop, S.; Tardon, A.; Fernandez, M. F.; Freire, C.; Goni, F.; Basagana, X.; Kogevinas, M.; Grimalt, J. O.; Sunyer, J.; Project, I. Exposure to trihalomethanes through different water uses and birth weight, small for gestational age, and preterm delivery in Spain. *Environ. Health Perspect.* **2011**, *119* (12), 1824−1830.
(32) Aurrekoetxea, J. J.; Murcia, M.; Rebagliato, M.; Lopez, M. J.; Castilla, A. M.; Santa-Marina, L.; Guxens, M.; Fernandez-Somoano, A.; Espada, M.; Lertxundi, A.; Tardon, A.; Ballester, F. 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* **2013**, *3* (1), e002034.
(33) Llop, S.; Casas, L.; Santa Marina, L.; Estarlich, M.; Fernandez-Somoano, A.; Esplugues, A.; Jimenez, A.; Zock, J. P.; Tardon, A.; Marco, A.; Ballester, F. Prenatal and postnatal residential usage of insecticides in a multicenter birth cohort in Spain. *Sci. Total Environ.* **2013**, *445−446*, 273−280.
(34) Dadvand, P.; de Nazelle, A.; Figueras, F.; Basagana, X.; Su, J.; Amoly, E.; Jerrett, M.; Vrijheid, M.; Sunyer, J.; Nieuwenhuijsen, M. J. Green space, health inequality and pregnancy. *Environ. Int.* **2012**, *40*, 110−5.
(35) European Commission Working Group Assessment of Exposure to Noise (WG-AEN). *Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise Exposure*; European Commission, 2006.
(36) ICGC. Cartographic base (topographical base of Catalonia 1:5.000) In Cartogràfic i Geològic de Catalunya.
(37) CNIG, CartoCiudad In 1.4 ed.; Instituto Geográfico Nacional de España, 2014.
(38) Walawender, J. P.; Hajto, M. J.; Iwaniuk, P. In A new ArcGIS toolset for automated mapping of land surface temperature with the use of LANDSAT satellite data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, July 22−27, 2012; pp 4371−4374.
(39) Gascon, M.; Verner, M. A.; Guxens, M.; Grimalt, J. O.; Forns, J.; Ibarluzea, J.; Lertxundi, N.; Ballester, F.; Llop, S.; Haddad, S.; Sunyer, J.; Vrijheid, M. Evaluating the neurotoxic effects of lactational exposure to persistent organic pollutants (POPs) in Spanish children. *NeuroToxicology* **2013**, *34*, 9−15.
(40) Baccarelli, A.; Pfeiffer, R.; Consonni, D.; Pesatori, A. C.; Bonzini, M.; Patterson, D. G., Jr; Bertazzi, P. A.; Landi, M. T. Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. *Chemosphere* **2005**, *60* (7), 898−906.
(41) Lenters, V.; Portengen, L.; Smit, L. A.; Jonsson, B. A.; Giwercman, A.; Rylander, L.; Lindh, C. H.; Spano, M.; Pedersen, H. S.; Ludwicki, J. K.; Chumak, L.; Piersma, A. H.; Toft, G.; Bonde, J. P.; Heederik, D.; Vermeulen, R. Phthalates, perfluoroalkyl acids, metals and organochlorines and reproductive function: a multipollutant assessment in Greenlandic, Polish and Ukrainian men. *Occup. Environ. Med.* **2015**, *72* (6), 385−93.
(42) Dominici, F.; Peng, R. D.; Barr, C. D.; Bell, M. L. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. *Epidemiology* **2010**, *21* (2), 187−94.
(43) Dadvand, P.; Ostro, B.; Figueras, F.; Foraster, M.; Basagana, X.; Valentin, A.; Martinez, D.; Beelen, R.; Cirach, M.; Hoek, G.; Jerrett, M.; Brunekreef, B.; Nieuwenhuijsen, M. J. Residential proximity to major roads and term low birth weight: the roles of air pollution, heat, noise, and road-adjacent trees. *Epidemiology* **2014**, *25* (4), 518−25.
(44) Foraster, M.; Kunzli, N.; Aguilera, I.; Rivera, M.; Agis, D.; Vila, J.; Bouso, L.; Deltell, A.; Marrugat, J.; Ramos, R.; Sunyer, J.; Elosua, R.; Basagana, X. High blood pressure and long-term exposure to indoor noise and air pollution from road traffic. *Environ. Health Perspect.* **2014**, *122* (11), 1193−1200.
(45) Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **2004**, *74* (4), 765−9.
(46) Gao, X. Multiple testing corrections for imputed SNPs. *Genet Epidemiol* **2011**, *35* (3), 154−8.
(47) Chadeau-Hyam, M.; Campanella, G.; Jombart, T.; Bottolo, L.; Portengen, L.; Vineis, P.; Liquet, B.; Vermeulen, R. C. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ. Mol. Mutagen* **2013**, *54* (7), 542−57.
(48) Agay-Shay, K.; Martinez, D.; Valvi, D.; Garcia-Esteban, R.; Basagana, X.; Robinson, O.; Casas, M.; Sunyer, J.; Vrijheid, M. Exposure to Endocrine-Disrupting Chemicals during Pregnancy and Weight at 7 Years of Age: A Multi-pollutant Approach. *Environ. Health Perspect.* **2015**, DOI: 10.1289/ehp.1409049.

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641

(49) White, I. R.; Royston, P.; Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* **2011**, *30* (4), 377−99.
(50) Al-Kandari, N. M.; Jolliffe, I. T. Variable selection and interpretation in correlation principal components. *Environmetrics* **2005**, *16* (6), 659−672.
(51) Nieuwenhuijsen, M. J.; Donaire-Gonzalez, D.; Rivas, I.; de Castro, M.; Cirach, M.; Hoek, G.; Seto, E.; Jerrett, M.; Sunyer, J. Variability in and Agreement between Modeled and Personal Continuously Measured Black Carbon Levels Using Novel Smartphone and Sensor Technologies. *Environ. Sci. Technol.* **2015**, *49* (5), 2977−2982.
(52) Braun, J. M.; Kalkbrenner, A. E.; Calafat, A. M.; Bernert, J. T.; Ye, X.; Silva, M. J.; Barr, D. B.; Sathyanarayana, S.; Lanphear, B. P. Variability and predictors of urinary bisphenol A concentrations during pregnancy. *Environ. Health Perspect.* **2011**, *119* (1), 131−137.
(53) Winquist, A.; Kirrane, E.; Klein, M.; Strickland, M.; Darrow, L. A.; Sarnat, S. E.; Gass, K.; Mulholland, J.; Russell, A.; Tolbert, P. Joint effects of ambient air pollutants on pediatric asthma emergency department visits in Atlanta, 1998−2004. *Epidemiology* **2014**, *25* (5), 666−73.
(54) Wild, C. P. The exposome: from concept to utility. *Int. J. Epidemiol* **2012**, *41* (1), 24−32.
(55) Vrijheid, M.; Slama, R.; Robinson, O.; Chatzi, L.; Coen, M.; van den Hazel, P.; Thomsen, C.; Wright, J.; Athersuch, T. J.; Avellana, N.; Basagana, X.; Brochot, C.; Bucchini, L.; Bustamante, M.; Carracedo, A.; Casas, M.; Estivill, X.; Fairley, L.; van Gent, D.; Gonzalez, J. R.; Granum, B.; Grazuleviciene, R.; Gutzkow, K. B.; Julvez, J.; Keun, H. C.; Kogevinas, M.; McEachan, R. R.; Meltzer, H. M.; Sabido, E.; Schwarze, P. E.; Siroux, V.; Sunyer, J.; Want, E. J.; Zeman, F.; Nieuwenhuijsen, M. J. The human early-life exposome (HELIX): project rationale and design. *Environ. Health Perspect.* **2014**, *122* (6), 535−44.

DOI: 10.1021/acs.est.5b01782
Environ. Sci. Technol. 2015, 49, 10632−10641