A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP4858.

Organophosphate Pesticide Exposure in Pregnancy in Association with Ultrasound and Delivery Measures of Fetal Growth

Kelly K. Ferguson,[1,2] Michiel A. van den Dries,[2,3] Romy Gaillard,[2,4] Anjoeka Pronk,[5] Suzanne Spaan,[5] Henning Tiemeier,[3,6] and Vincent W. V. Jaddoe[2,4,7]

[1]Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA
[2]Generation R Study Group, Erasmus University Medical Center MC (Erasmus MC), Rotterdam, Netherlands
[3]Department of Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, Netherlands
[4]Department of Pediatrics, Erasmus MC, Rotterdam, Netherlands
[5]Department of Risk Analysis for Products in Development, Netherlands Organisation for Applied Scientific Research (TNO), Zeist, Netherlands
[6]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[7]Department of Epidemiology, Erasmus MC, Rotterdam, Netherlands

Background: Perturbations in fetal growth may have adverse consequences for childhood and later life health. Organophosphate pesticide (OP) exposure has been associated with reduced birth weight at delivery but results are not consistent. We investigated this question by utilizing ultrasound measures of size *in utero* in combination with measures from delivery.

Methods: Within Generation R, a population-based prospective cohort conducted between 2002 and 2006 in Rotterdam, Netherlands, we measured dialkyl phosphates (DAPs), OP metabolites, in urine samples from early, middle, and late pregnancy and created a subject-specific average to estimate OP exposure ($n = 784$). Ultrasound measures of head circumference, femur length, and estimated fetal weight from middle and late pregnancy and delivery measures were converted to standard deviation scores (SDS). Associations with DAP average were examined in linear mixed effects models that included an interaction term between gestational age at measurement and DAP average to investigate whether the relationship differed over time. Windows of vulnerability to exposure were assessed by modeling urinary DAPs from each visit in relation to growth measurements.

Results: A 10-fold increase in average DAPs was associated with a -0.53 SDS decrease in fetal length (95% CI $= -0.83, -0.23$) and a -0.32 SDS decrease in estimated fetal weight (95% CI $= -0.59, -0.04$) at 20 weeks of gestation. These differences corresponded to 5% and 6% decreases relative to the mean. Effect estimates were greatest in magnitude for DAP concentrations measured early in pregnancy. Associations between average DAPs and growth measures at delivery were positive but not significant for head circumference and length and were null for weight.

Conclusions: Maternal urinary DAPs were associated with decreased fetal weight and length measured during mid-pregnancy, but not at delivery. https://doi.org/10.1289/EHP4858

Introduction

Perturbations in normal fetal growth are linked to numerous adverse health outcomes both in childhood (Miller et al. 2016) and later life (Barker 2006). Suboptimal fetal growth is classically approximated by birth weight at delivery with or without adjustment for gestational duration. However, for diagnostic purposes, assessment of fetal growth longitudinally during pregnancy is preferred (Resnik 2018). In research, repeated ultrasound measures of growth allow for the *a*) improved ability to detect deviations from normality that occur during gestation, not just at delivery; *b*) investigation of rates of change in growth, rather than a snapshot of size; and *c*) assessment of specific fetal growth measures, such as length as an indicator of skeletal size, which are not fully captured by birth weight alone. Utilizing these data, researchers have demonstrated specific time periods in pregnancy where changes in rate of growth may have the greatest impact on childhood health outcomes (e.g., adiposity, neurodevelopment) (Gishti et al. 2014; Henrichs et al. 2010). Similarly, studies of environmental factors and fetal growth have used these data to augment understanding of windows when exposures have the strongest influence on growth and which specific anthropometric parameters are most affected (e.g., head circumference vs. weight) (Aguilera et al. 2010; Ferguson et al. 2016; Philippat et al. 2014; Snijder et al. 2013).

To our knowledge, longitudinal ultrasounds in pregnancy have not been used to investigate the association between organophosphate pesticide (OP) exposure and fetal growth. OPs such as dimethoate and parathion are a class of high-production insecticides with neurotoxic capacity. Exposure can occur through occupational use or proximity to areas with agricultural application, but most populations are exposed through diet (Llop et al. 2017; Lu et al. 2008; Sokoloff et al. 2016; van den Dries et al. 2018). There is strong biologic plausibility for an effect of OPs on *in utero* growth and development through interference with adenyl cyclase activity, which is crucial for cell differentiation (Song et al. 1997); disruption of normal thyroid hormone function in the mother or fetus (Campos and Freire 2016); or dysregulation of nutrient transport across the placenta (Eskenazi et al. 1999). Evidence for an association with birth weight has been demonstrated in some but not all rodent studies of OPs (Breslin et al. 1996; Chanda and Pope 1996; Maurissen et al. 2000; Muto et al. 1992). Results from human studies on the association between biomarkers of OPs and birth weight, including a recent pooled analysis, have also been ambiguous, and associations may differ by individuals' ability to detoxify OPs by the paraoxonase enzyme (Harley et al. 2016).

In the present study, we investigated the association between maternal OP exposure in pregnancy and fetal growth as assessed by repeated ultrasound measurements during pregnancy in combination with neonatal assessments. We utilized urinary dialkyl phosphates (DAPs), metabolites of OPs, measured in urine samples collected at three time points in pregnancy as proxies of exposure. Our primary aim was to assess associations of average DAPs over pregnancy with repeated measures of head circumference, length, and weight measured at two time points during

Address correspondence to Kelly K. Ferguson, 111 T.W. Alexander Dr., P.O. Box 12233, MD A3-05, Research Triangle Park, NC 27709 USA. Telephone: (984) 287-3700. Email: Kelly.ferguson2@nih.gov

Supplemental Material is available online (https://doi.org/10.1289/EHP4858).
The authors declare that they have no actual or potential conflicts of interest.
Received 10 December 2018; Revised 22 July 2019; Accepted 25 July 2019; Published 16 August 2019.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

pregnancy by ultrasound and at delivery. Our secondary aim was to identify potential windows of vulnerability to exposure by examining outcomes in association with DAP concentrations at each individual time point. We additionally examined effect modification of these associations by fetal sex and *PON1* genotype.

Methods

Study Population

Generation R is a prospective population-based birth cohort designed to identify early environmental and genetic determinants of development throughout life and which has been described in detail previously (Kooijman et al. 2016). Briefly, all mothers who resided in the study area in Rotterdam, Netherlands, and had a delivery date between April 2002 and January 2006 were eligible. Mothers were enrolled during pregnancy or in the first months after the birth of their child when newborns visited the routine child health centers. Among the 9,778 mothers who participated in the study 8,879 (91%) were enrolled during pregnancy. Among the 4,918 women enrolled during pregnancy between February 2004 and January 2006, spot urine specimens during early, middle, and late pregnancy (<18, 18–25, and >25 weeks of gestational age, respectively) were collected at the time of routine ultrasound examinations. In total, 2,083 women provided a complete set of three urine specimens. The study protocol underwent human subjects review at Erasmus Medical Center, Rotterdam, Netherlands (institutional review board registration no. IRB00001482, MEC 198.782.2001.31). Mothers provided written informed consent for themselves and their children.

Among the women with urine specimens collected at each of the three visits in pregnancy, 1,449 had complete information on childhood health assessments (Kooijman et al. 2016). From these women, 800 were randomly selected for a study designed to assess the relationship between prenatal exposure to OPs and childhood neurodevelopmental outcomes (van den Dries et al. 2018). Due to limitations in urine sample volume, 784 individuals were included in the final study population ($n = 778$ with three samples; $n = 5$ with two samples; $n = 1$ with one sample). A flow chart describing the selection process is shown in Figure S1. Women in this subset had higher education levels and were slightly older and a greater proportion were Dutch compared with the broader Generation R cohort (Kooijman et al. 2016).

Ultrasound and Delivery Measures of Size

During pregnancy, ultrasound scans were performed to calculate gestational age and to measure fetal growth on the entire study population, as described in detail elsewhere (Gaillard et al. 2014). Head circumference and length were measured in middle and late pregnancy and estimated fetal weight for each time point was calculated using the formula of Hadlock et al. (1985). At birth, head circumference, length, and weight were measured. Standard deviation scores (SDS) for each measurement were calculated using longitudinal growth curves that accounted for gestational age at measurement but not fetal sex (Gaillard et al. 2014).

Urinary Dialkyl Phosphate Measurement

At each of the three study visits, urine samples were collected from participants in polypropylene cups and stored until analysis at −20°C (Jusko et al. 2019; Kruithof et al. 2014). Six nonspecific DAPs were measured using gas chromatography coupled with tandem mass spectrometry (GC-MS/MS) at the Institute National de Santé Publique (INSPQ) in Quebec, Canada, with methods described in detail elsewhere (Haines and Murray 2012; Jusko et al. 2019). These measurements included three dimethyl metabolites (dimethylphosphate, dimethythiophosphate, and dimethyldithiophosphate; DMPs) and three diethyl metabolites (diethylphosphate, diethylthiophosphate, and diethyldithiophosphate; DEPs). Limits of detection were between 0.06 and 0.5 μg/L and coefficients of variation for inter-day reliability were <10% (van den Dries et al. 2018). Values below the limit of detection were imputed with machine-reported values when available.

We calculated nanomolar sums of DMPs, DEPs, and total DAPs using molecular weights (Jusko et al. 2019; van den Dries et al. 2018). To adjust for urine dilution, we measured creatinine concentrations using the Jaffe reaction and corrected each sum so that final concentrations are presented in nanomoles per gram creatinine. Finally, we calculated subject-specific geometric averages of DMP, DEP, and DAP concentrations from levels measured at each of the three visits in pregnancy in order to create more stable estimates of exposure for our primary aim (van den Dries et al. 2018).

PON1 genotyping

Cord blood from 523 children included in the present analysis was genotyped using Illumina 610K and 660W arrays, as described previously (Jusko et al. 2019). We examined single nucleotide polymorphisms (SNPs) for $PON1_{-Q192}$, which was directly genotyped, and four other SNPs that were imputed in the genotype data set, including rs705379 ($PON1_{-108}$), rs705381 ($PON1_{-161}$), rs854560 ($PON1_{-L55M}$), and rs854572 ($PON1_{-909}$). MaCH 1.0 (Li et al. 2010) was used to impute the 1000 Genomes Iv3 reference panel (1000 Genomes Project Consortium et al. 2015), and all four imputed SNPs had excellent imputation quality ($R^2 > 0.95$) and high minor allele frequencies (>26%). For individuals with no genotyping data available, all SNPs were imputed as described below.

Statistical Methods

All analyses were performed using R (version 3.4.3; R Development Core Team). To address missing data, we imputed the data set 10 times using multiple imputation by chained equation (MICE) in R (package *mice*) (van Buuren and Groothuis-Oudshoorn 2010). For DAPs, a small number of concentrations were missing due to insufficient sample or machine error (≤5 measurements for any visit for DMPs; ≤23 for DEPs; ≤5 for creatinine). Imputations were performed prior to calculating nanomolar sums, creatinine correction, and calculation of subject-specific averages. Missing covariates listed in Table 1 (<20% for all) were imputed, and the following covariates were additionally included as predictors for imputation: maternal education level; caloric intake; caloric intake from vegetables and caloric intake from fruits; paternal education level; maternal ethnicity; and body mass index (BMI). We also imputed missing ultrasound and delivery SDS of fetal or newborn size. Due to the correlation between size measurements, only birth-weight SDS was included as a predictor in the MICE procedure. Finally, for individuals missing all *PON1* genotyping data, we imputed SNPs in the same MICE procedure as has been done in our previous analyses; however, *PON1* was not used as a predictor in the imputation step due to a high proportion of missing measures. Thus, unless stated otherwise, all models presented contain the full sample ($n = 784$) and complete observations at all time points.

We calculated distributions of demographic characteristics and DAP averages and examined Pearson correlations between DAPs at individual time points and for averages. We calculated raw (i.e., unstandardized) ultrasound and delivery measures of size in the unimputed data set for interpretation purposes. To address our primary research question, we created linear mixed effects models using the *nlme* package (Pinheiro et al. 2014), modeling average

Table 1. Demographic and lifestyle characteristics in a subset ($n = 784$) of mothers with singleton live births from the Generation R Study population.

Characteristic	Median (25th, 75th) or n (%)
Maternal age (y)	31 (28, 34)
<20	14 (1.79)
20–< 25	79 (10.1)
25–< 30	208 (26.5)
30–< 35	360 (45.9)
≥35	123 (15.7)
Maternal ethnicity	
Dutch	451 (57.5)
Other Western	70 (8.93)
Non-Western	263 (33.6)
Maternal education	
Low	113 (14.9)
Intermediate	229 (30.2)
High	417 (54.9)
Missing	25 (3.12)
Household income (Euros/month)	
<1,200	86 (11.0)
1,200–2,000	113 (14.4)
>2,000	483 (61.6)
Missing	102 (13.0)
Marital status	
Partner	677 (89.7)
No partner	78 (10.3)
Missing	29 (3.70)
Weight prepregnancy (kg)	64.0 (58.0, 72.0)
Missing	96 (12.2)
Height at Visit 1 (cm)	168 (163, 173)
Missing	1 (0.13)
Parity	
0	486 (62.0)
1	208 (26.5)
≥2	86 (11.0)
Missing	4 (0.51)
Smoking	
No smoking during pregnancy	555 (70.8)
Until pregnancy recognized	64 (8.16)
Continued during pregnancy	102 (13.0)
Missing	63 (8.04)
Alcohol consumption	
No consumption during pregnancy	273 (34.8)
Until pregnancy recognized	130 (16.6)
Continued occasionally	293 (37.4)
Continued frequently	48 (6.12)
Missing	40 (5.10)
Folic acid intake	
None	98 (12.5)
Started in first 10 weeks of pregnancy	212 (27.0)
Started preconception	319 (40.7)
Missing	155 (19.8)
Fetal sex	
Male	398 (50.8)
Female	386 (49.2)
Preterm (<37 weeks of gestation)	
No	762 (97.1)
Yes	22 (2.90)
Low birth weight (<2,500 g)	
No	761 (97.0)
Yes	23 (3.0)

DMP, DEP, or DAP exposure over pregnancy in relation to repeated SDS of head circumference, length, or weight (ultrasound measures from middle and late pregnancy in combination with birth measurements at delivery). In this and other pregnant populations, within-individual variation in urinary DAP concentrations is greater than variation between individuals (Casas et al. 2018; Spaan et al. 2015). This reflects daily variation in exposure through, for example, variable dietary patterns, as well as rapid metabolic clearance of these compounds (van den Dries et al. 2018). Consequently, if exposures are relatively consistent over longer periods of time, average DAP concentrations based on multiple urine samples should provide a more accurate measure of usual exposure at any point during pregnancy than DAP concentrations measured in an individual sample. Therefore, in our primary analyses, we used pregnancy average DAP concentrations to estimate usual exposure across pregnancy despite the fact that this included urinary biomarkers from late pregnancy that were collected after the mid-pregnancy ultrasound. DAP averages were $\log_{10}$-transformed for analysis to improve model fit. All models included a random intercept for each subject as well as a random slope for gestational age at growth measurement. We additionally included an interaction term between DAP average and gestational age at growth measurement to allow the exposure–outcome association to differ by timing of outcome measurement. Because results differed based on the timing of outcome measurement and because presentation of main effect and interaction terms can be difficult to interpret, we presented results from models where the intercept was varied so that the results would represent associations where the outcome was measured at 20, 30, and 40 weeks of gestation.

For adjusted models, we included an *a priori* set of covariates based on previously observed associations with the exposure and outcome or covariates known to improve the precision of the outcome estimate. These included fetal sex (categorical), maternal age (continuous), prepregnancy weight (continuous), height (continuous), maternal education level (categorical), maternal ethnicity (categorical), parity (categorical), smoking in pregnancy (categorical), alcohol use in pregnancy (categorical), folic acid use, and gestational age at growth measurement (continuous). Folic acid was included in this study population because it has been associated with fetal growth and because it is a strong indicator of socioeconomic status (SES), a predictor of urinary DAP concentrations (Timmermans et al. 2008, 2009; van den Dries et al. 2018). Maternal height and weight were included instead of the aggregate BMI because each is an independent predictor of fetal growth (Gardosi et al. 1992).

Our second aim was to examine windows of vulnerability to exposure. To do so, we created cross-sectional models of total urinary DAP concentrations measured at each study visit in relation to outcome measurements at middle and late pregnancy and at delivery (i.e., one exposure time point and one outcome time point per model). For these analyses, we examined only outcomes at the same or subsequent visits (i.e., we did not model late pregnancy exposure biomarkers in association with middle pregnancy fetal growth measurements). All models retained the same covariates as those used in the repeated measures analyses. Results for all cross-sectional models are reported for associations with outcomes at early, middle, or late pregnancy and are referred to as such.

To assess effect modification of the relationship between DAPs and fetal growth by fetal sex, we examined associations in models stratified by this variable, and we additionally included a two-way interaction term between sex and exposure in repeated measures models to test the significance of any observed differences. A similar approach was used to estimate effect modification by *PON1* genotype. Interaction terms with $p < 0.05$ were considered statistically significant.

To test the robustness of our results, we first examined the influence of imputing outcome measurements. To do so, we created a new set of 10 imputed data sets in which the low proportion missing for exposures and covariates were imputed, but the outcome was not. We then recreated linear mixed effects models for comparison. Second, we examined results with DMPs and DEPs included in the same model in order to distinguish the effects of the two classes. Third, because these metabolites demonstrate only weak-to-moderate reliability over pregnancy (e.g., intraclass correlation coefficient for DAP metabolites = 0.30)

Table 2. Distribution of fetal and neonatal anthropometric parameters prior to imputation.

Anthropometric parameter	*n* (%)	Mean (SD)
Middle pregnancy (ultrasound)		
Gestational age (weeks)	784 (100)	20.4 (0.92)
Head circumference (mm)	774 (98.7)	178 (12.3)
Length (mm)	779 (99.4)	33.1 (2.98)
Estimated fetal weight (g)	777 (99.1)	369 (73.8)
Late pregnancy (ultrasound)		
Gestational age (weeks)	784 (100)	30.4 (0.83)
Head circumference (mm)	777 (99.1)	286 (11.4)
Length (mm)	784 (100)	57.6 (2.80)
Estimated fetal weight (g)	782 (99.7)	1,626 (238)
Delivery (physical exam)		
Gestational age (weeks)	784 (100)	40.1 (1.48)
Birth head circumference (cm)	478 (61)	33.7 (1.72)
Birth length (cm)	562 (71.7)	50.4 (2.24)
Birth weight (g)	784 (100)	3,452 (505)

(Spaan et al. 2015), we examined the effect of adjusting for measurement error by applying regression calibration (Hardin et al. 2003). We applied the calibration to repeated measures models of average DAP concentrations in associations with each fetal growth outcome. Fourth, to test the robustness of our results to adjustment for additional SES factors, we examined models additionally adjusted for marital status and income level. Fifth, because season has been associated both with urinary DAP (van den Dries et al. 2018) and with birth weight in some studies (Strand et al. 2011), we examined the effect of additionally adjusting for this potential confounder. Finally, we examined associations after removing babies who were born preterm (i.e., prior to 37 weeks of completed gestation) in order to determine whether or not our results could be attributed to gestational age at delivery rather than size.

Results

Of the 784 women included in the present analysis, the median maternal age was 31 y, most of the women were Dutch (58%), and the prepregnancy median weight was 64 kg (Table 1). For 62% of women this was their first pregnancy, and smoking and alcohol use was low to moderate in the study population. A small percentage of women never took folic acid supplements either prior to or at any point during pregnancy.

The median gestational ages for middle and late pregnancy ultrasounds were 20.4 weeks [95% confidence interval (CI): 20.3, 20.5] and 30.4 weeks (95% CI: 30.3, 30.5), respectively, and the median gestational age at delivery was 40 weeks (95% CI: 40.0, 40.2) (Table 2). Almost all participants included in the present analysis had ultrasound measurements available at these two time points prior to imputation. All participants had data available on birth weight at delivery, but a smaller proportion had head circumference (61%) or birth length (72%) assessed. Thus, a larger proportion of head circumference and birth length measurements were imputed. Distributions of urinary DMPs, DEPs, and DAPs by study visit and on average are presented in Table 3. As previously reported, concentrations by visit showed weak-to-moderate reliability (intraclass correlation coefficients 0.14–0.38) (Spaan et al. 2015). DMPs and DAPs were highly correlated both for averages and at individual study visits (Pearson $r = 0.97$–0.98), but DEPs were less correlated with DAPs (Pearson $r = 0.53$–0.63) and DMPs (Pearson $r = 0.43$–0.47) (see Table S1).

Primary Analysis: Repeated Measures Models

Effect estimates from fully adjusted repeated measures models, accounting for interaction between exposure and gestational age at growth measurement, demonstrated that associations between pregnancy averages of exposure and outcomes differed based on the timing of outcome measurement (i.e., interactions between exposure and time were statistically significant; see Table S2). For the presentation of results, we calculated effect estimates for outcomes at 20, 30, and 40 weeks (Table 4). At 20 weeks, a 10-fold increase in pregnancy-averaged total DAPs was associated with a 0.53-SDS shorter length (95% CI: −0.83, −0.23) and a 0.32-SDS lower weight (95% CI: −0.59, −0.04). For length, this difference corresponds to −2 mm, or −5%, relative to the mean for length at 20 weeks of gestation. For weight, this corresponds to −24 g, or −6%, relative to the mean for estimated fetal weight at 20 weeks of gestation. DMPs and DEPs individually were also inversely associated with length and weight at this time point, but none of the DAPs were significantly associated with differences in head circumference.

Table 3. Distribution of urinary organophosphate pesticide metabolite concentrations by study visit and on average (nmol/g creatinine) in a subset of the Generation R Study population ($n = 784$).

		Percentiles			
Metabolites/anthropometric parameter	Geometric mean	25th	50th	75th	95th
Total dimethyl phosphates (DMPs)[a]					
Early pregnancy	249.8	148.6	244.1	413.6	860.0
Middle pregnancy	263.6	168.8	268.6	415.4	854.3
Late pregnancy	247.4	157.1	247.8	398.9	863.1
Average	253.5	183.1	259.1	355.3	582.9
Total diethyl phosphates (DEPs)[b]					
Early pregnancy	42.9	25.1	43.1	79.3	175.6
Middle pregnancy	40.5	23.2	41.6	74.3	179.8
Late pregnancy	40.1	21.6	41.5	77.3	176.3
Average	40.0	28.2	42.6	64.6	116.6
Total dialkyl phosphates (DAPs)[c]					
Early pregnancy	308.4	188.1	306.9	499.3	989.0
Middle pregnancy	317.9	206.7	316.5	485.9	1,001.9
Late pregnancy	301.5	194.0	307.9	489.0	984.8
Average	309.2	226.4	311.0	438.8	687.3

Note: Number of metabolites missing due to machine error in early, middle, and late pregnancy: DMDTP: 0, 0, 3; DMP: 0, 0, 0; DMTP: 0, 0, 1; DEDTP: 0, 1, 1; DEP: 1, 0, 0; DETP: 13, 22, 16. Number of metabolites missing due to insufficient urine volume for analyses: early pregnancy = 5; middle pregnancy = 1; late pregnancy = 1 (6 participants total). Values in this table include imputed data.

[a]DMPs represent a molar sum of dimethylphosphate (DMP), dimethylthiophosphate (DMTP), and dimethyldithiophosphate (DMDTP).

[b]DEPs represent a molar sum of diethylphosphate (DEP), diethylthiophosphate (DETP), and diethyldithiophosphate (DEDTP).

[c]DAPs represent a molar sum of all of the above. Percent below the limit of detection by individual metabolite in early, middle, and late pregnancy: DMDTP: 19.9, 18.1, 18.0; DMP: 0.1, 0, 0; DMTP: 3.5, 3.6, 2.4; DEDTP: 81.1, 84.5, 85.0; DEP: 2.7, 5.4, 4.1; DETP: 12.1, 11.8, 11.7.

Table 4. Adjusted difference in fetal head circumference, length, or weight standard deviation score (SDS) at selected weeks of gestation in association with pregnancy average urinary organophosphate pesticide metabolite concentrations in a subset of the Generation R Study population ($n = 784$).

	Difference in SDS score (95% CI)		
Anthropometric parameter/gestational age	Total DMPs[a]	Total DEPs[b]	Total DAPs[c]
Head circumference (weeks)			
20	−0.11 (−0.42, 0.19)	0.00 (−0.13, 0.14)	−0.09 (−0.40, 0.23)
30	0.12 (−0.14, 0.38)	0.07 (−0.04, 0.18)	0.13 (−0.14, 0.39)
40	0.35 (−0.07, 0.78)	0.13 (−0.06, 0.33)	0.34 (−0.11, 0.79)
Length (weeks)			
20	−0.46 (−0.75, −0.17)	−0.17 (−0.30, −0.04)	−0.53 (−0.83, −0.23)
30	−0.09 (−0.32, 0.14)	−0.06 (−0.17, 0.04)	−0.13 (−0.37, 0.10)
40	0.29 (−0.05, 0.63)	0.05 (−0.11, 0.21)	0.27 (−0.08, 0.61)
Weight (weeks)			
20	−0.27 (−0.54, 0.00)	−0.16 (−0.28, −0.04)	−0.32 (−0.59, −0.04)
30	−0.10 (−0.32, 0.12)	−0.08 (−0.18. 0.02)	−0.13 (−0.36, 0.09)
40	0.06 (−0.22, 0.35)	0.00 (−0.13, 0.13)	0.05 (−0.25, 0.34)

Note: Results from linear mixed effects models adjusted for fetal sex, maternal age (continuous), prepregnancy weight (continuous), height (continuous), education level (categorical), maternal ethnicity (categorical), parity (categorical), smoking (categorical), alcohol use (categorical), folic acid use (categorical), and gestational age at ultrasound or delivery (continuous). Models contain an interaction term between exposure concentration and gestational age at ultrasound/delivery, a random intercept for each participant, and a random slope for gestational age at ultrasound/delivery. Values in this table include imputed data. CI, confidence interval; DAPs, dialkyl phosphates; DEPs, diethyl phosphates; DMPs, dimethyl phosphates.
[a]DMPs represent a molar sum of dimethylphosphate (DMP), dimethylthiophosphate (DMTP), and dimethyldithiophosphate (DMDTP).
[b]DEPs represent a molar sum of diethylphosphate (DEP), diethylthiophosphate (DETP), and diethyldithiophosphate (DEDTP).
[c]DAPs represent a molar sum of all of the above.

Interaction terms between exposure and gestational age indicated that associations became weaker as pregnancy progressed, so that at 30 or 40 weeks no significant associations between DAPs and fetal measurements were observed (Table 4). To illustrate these effects, we plotted estimated coefficients and confidence intervals by time for associations between DAPs and head circumference (Figure 1A), length (Figure 1B), and weight (Figure 1C). This shows that at delivery associations were null for weight and positive but nonsignificant for head circumference and length. These results are also consistent with those from cross-sectional models of pregnancy averages in association with growth measurements from each study visit (middle pregnancy, late pregnancy, and delivery). Cross-sectional associations with pregnancy averages are displayed in Figure 2D, with effect estimates in Table S3.

Secondary Analysis: Windows of Vulnerability to Exposure

Cross-sectional models of visit-specific urinary DAP concentrations in association with each outcome demonstrated some differences by timing of exposure (Figure 2A–C; see also Tables S4–S6). In general, total urinary DAPs measured in early pregnancy showed the strongest associations with length and weight (Figure 2A; Table S4). A 10-fold increase in concentrations measured in samples collected in early pregnancy was associated with lower fetal length ($\beta = -0.30$, 95% CI: −0.50, −0.10) and weight ($\beta = -0.22$, 95% CI: −0.4, −0.04) at mid-pregnancy, and also with lower fetal length ($\beta = -0.16$, 95% CI: −0.36, 0.04) and weight ($\beta = -0.19$, 95% CI: −0.39, 0.01) in late pregnancy, although the latter association was not statistically significant. Levels measured in mid-pregnancy samples were associated with lower fetal length ($\beta = -0.2$, 95% CI: −0.42, 0.02) and weight ($\beta = -0.14$, 95% CI: −0.34, 0.06) in mid-pregnancy but not in late pregnancy or at delivery (Figure 2B; Table S5). Finally, levels measured in late pregnancy samples were not associated with differences in length or weight but were positively associated with head circumference measured at delivery (Figure 2C; Table S6). Patterns for DMPs and DEPs were similar to the overall DAPs (see Figures S2–S3 and Tables S4–S6).

Effect Modification by Sex and Genotype

Interaction terms between total DAPs and sex demonstrated associations with length and weight that were stronger (i.e., more negative) for males compared with females (Table 5), although associations were still observed for females in models of length. Adjusted associations by gestational age and sex are displayed in Figures 3A–C for length, weight, and head circumference respectively. These figures show how the exposure–response associations differ in males and females based on timing of outcome measurement. Interaction terms between total DAPs and *PON1* genotype were not statistically significant ($p = 0.18$–0.93) and there were no clear differences in exposure–outcome associations according to SNP genotypes (see Tables S7A–E).

Sensitivity Analyses

For repeated measures models, associations without adjustment for covariates were slightly greater in magnitude (i.e., more negative) but otherwise similar to those observed in our primary analysis (see Table S8). Associations in models without imputed outcome were also similar, although effect estimates for head circumference were closer to the null (see Table S9). Summed DEPs and DMPs were moderately correlated (Pearson $r = 0.45$). When exposures were both included in a mutually adjusted model, associations were similar to the primary single-pollutant findings (see Table S10). In both, effect estimates were greater in magnitude (i.e., more negative) for DMPs compared with DEPs for all growth parameters. We additionally examined the effect of adjusting for the measurement error inherent in our exposure biomarkers by creating our repeated measures models of average urinary DAPs with application of regression calibration. As expected, we observed that the effect estimates were less attenuated (i.e., farther from the null) compared with the results from the primary model but that variances were increased (see Table S11). Finally, we observed minimal differences when we added marital status and family income as additional measures of SES to the model, when we included season of sample collection, or when our population was restricted to babies born at term (see Table S12–S14, respectively, for DAPs only).

Discussion

In a study of pregnant women in the Netherlands, we observed that elevated urinary biomarkers of OP exposure were associated with reduced fetal size as indicated by measurements of length and weight in mid-pregnancy, but not at delivery. Furthermore, elevated biomarkers of exposure from early as well as mid-pregnancy demonstrated associations with growth measurements that were greatest in magnitude and the most precise. The latter suggests that early

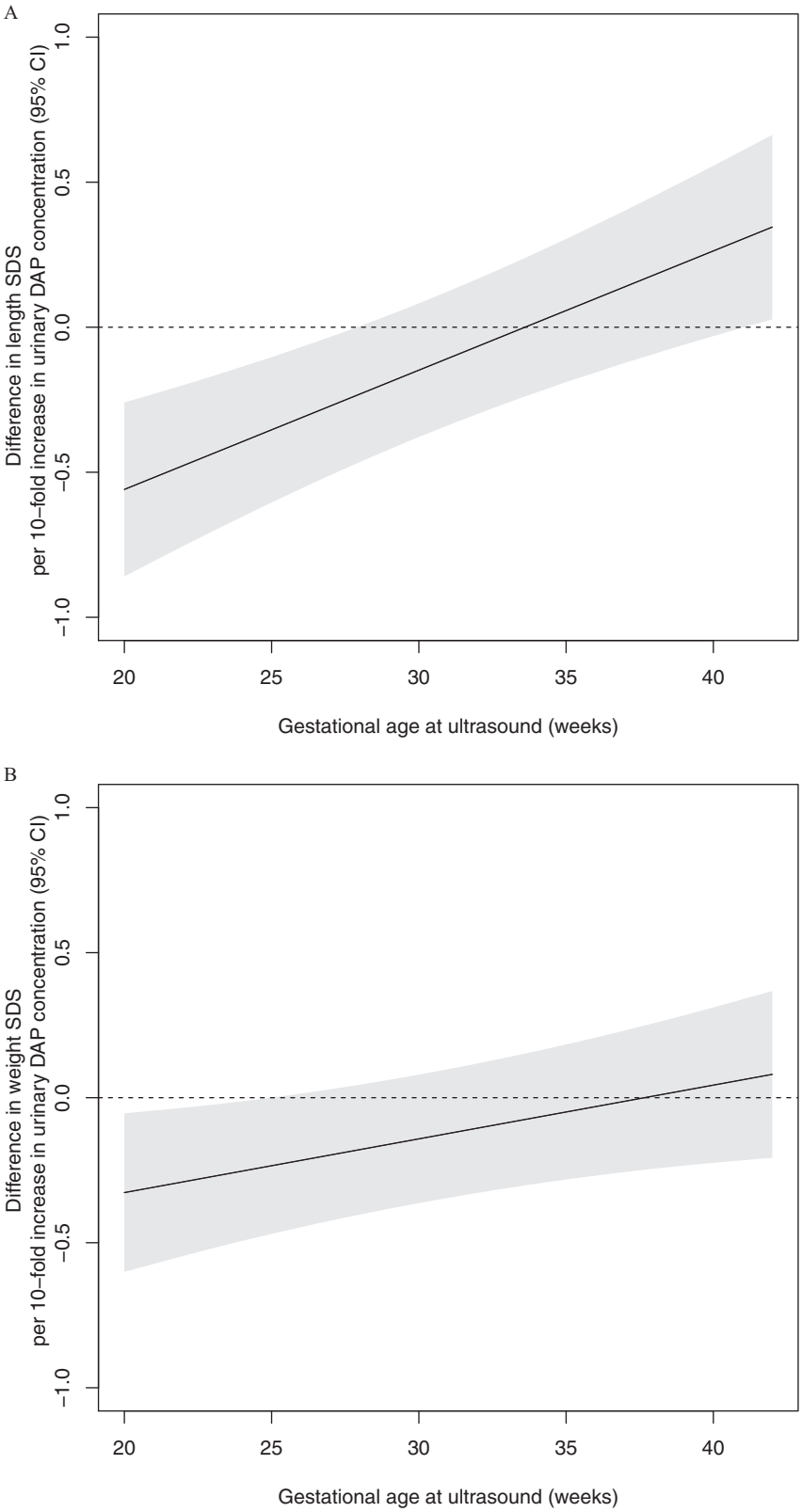

Figure 1. Adjusted repeated measures associations between pregnancy average total dialkyl phosphate (DAP) concentrations and standard deviation scores (SDS) of (A) head circumference, (B) length, and (C) weight by gestational age at growth measurement in the Generation R Study population ($n = 784$). Models adjusted for fetal sex, maternal age (continuous), prepregnancy weight (continuous), height (continuous), education level (categorical), maternal ethnicity (categorical), parity (categorical), smoking (categorical), alcohol use (categorical), folic acid use (categorical), and gestational age at ultrasound or delivery (continuous). Model contains an interaction term between exposure concentration and gestational age at ultrasound/delivery, a random intercept for each participant, and a random slope for gestational age at ultrasound/delivery. Main effect and interaction terms (95% CIs) for each plot are as follows: (A) −0.51 (−1.32, 0.29); 0.02 (−0.01, 0.05); (B) −1.33 (−2.00, −0.66); 0.04 (0.02, 0.06); (C) −0.68 (−1.24, −0.12); 0.02 (0.00, 0.04). This figure includes imputed data. Note: CI, confidence interval; DAPs, dialkyl phosphates; SDS, standard deviation scores.

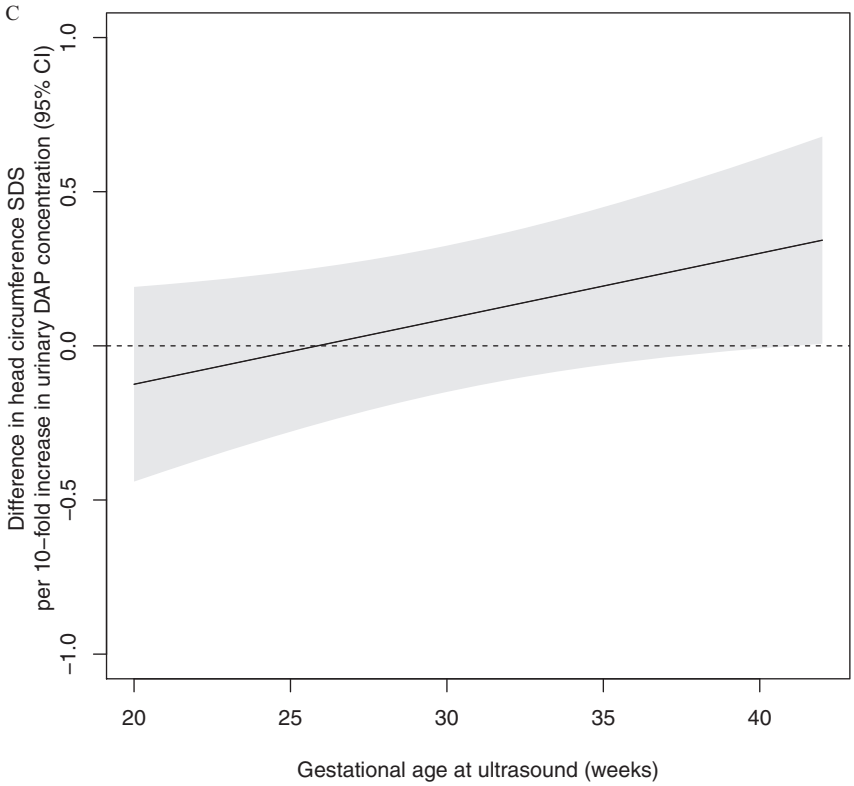

Figure 1. (*Continued.*)

pregnancy exposure may be an important vulnerable window for the relationship between OP exposure and fetal growth.

The results from our analysis may shed light on previous epidemiologic studies with ambiguous findings on this relationship. All previous studies have examined exposure in association with measurements at delivery, and few have found evidence of effects. This is consistent with what we would expect based on our study given that we observed no associations between exposure biomarkers and outcomes measured at delivery. A recent pooled study combined data from four U.S. studies for a powerful assessment of this research question (total sample size $\sim$ 1,100); however, no association was observed overall between prenatal urinary exposure biomarkers and birth weight, length, or head circumference (Harley et al. 2016). Other individual studies using urinary biomarkers have noted some associations between these exposures and outcomes, with differences by timing of exposure assessment, *PON1* genotype or expression, and, in some instances, fetal sex; however, no clear patterns emerge upon review of the data (Dalsager et al. 2018; Eskenazi et al. 2004; Huang et al. 2017; Liu et al. 2016; Naksen et al. 2015; Rauch et al. 2012; Wang et al. 2012; Wolff et al. 2007; Woods et al. 2017).

Some notable differences exist between this previous work and our present study. First, all but two of the previous studies had significantly smaller sample sizes ($n \sim 50$–450). However, those with similar sample sizes had null findings. The largest study to date was the pooled analysis, although the authors of that study noted the difficulties in combining these data across populations with differing demographics and exposure levels ($n \sim 1{,}000$) (Harley et al. 2016). The second largest study from the Odense Child Cohort in Denmark ($n = 858$) measured urinary DAP concentrations at $\sim$28 weeks of gestation and was unable to detect associations with birth weight, length, or abdominal or head circumference at delivery (Dalsager et al. 2018).

Second, all but three studies assessed urine concentrations in a single spot urine sample collected during gestation or at delivery. Those with repeated measures had largely null findings, although they were also more limited in sample size. Woods et al. ($n = 272$) averaged measures from 16 and 26 weeks of gestation and did not detect associations with birth weight (Woods et al. 2017). Naksen et al. ($n = 52$) and Huang et al. ($n = 105$) modeled urinary concentrations from two visits during pregnancy and at delivery separately, with primarily null results (Huang et al. 2017; Naksen et al. 2015). OPs are metabolized quickly in the human body, and urinary DAP concentrations show only moderate stability over pregnancy (Huang et al. 2017; van den Dries et al. 2018); thus, the availability of repeated measures for estimating more stable subject-specific averages is an advantage in our study. Even though these averages included a urinary measurement (late pregnancy) that was taken after the time of some outcome measurements (ultrasound measures from mid-pregnancy), the average is the best choice for exposure assessment because individual measurements are highly variable over time due to variations in exposure sources (e.g., diet) in combination with rapid metabolic excretion (Casas et al. 2018; Spaan et al. 2015). Consequently, if we assume that exposures are generally consistent over pregnancy, the average measure will be the best estimate.

In addition, availability of repeated measurements enabled us to investigate windows of vulnerability to exposure during gestation. We observed that urinary DAP concentrations from early pregnancy were associated with reduced length and estimated fetal weight in mid- and late pregnancy. DAP levels from mid- and late pregnancy, however, were not associated with growth measurements. This could suggest that early pregnancy is a particularly sensitive window to exposure. Early to mid-pregnancy is a time of rapid placental development that could be mediating these effects.

A third major difference between our study and those previously published is in exposure biomarker levels, which differ

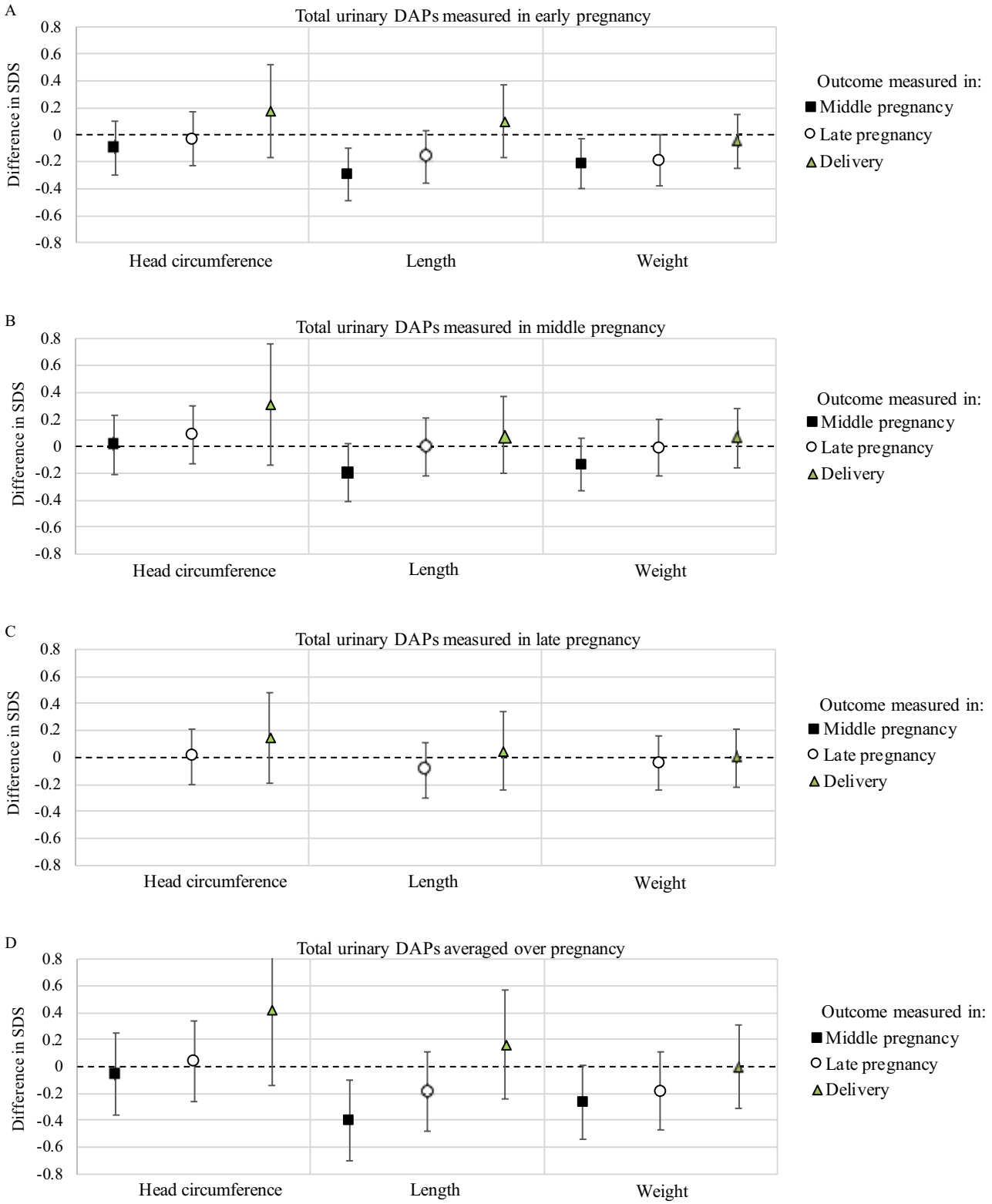

Figure 2. Adjusted cross-sectional associations between visit-specific total dialkyl phosphate (DAP) concentrations measured in total urinary DAPs in A) early pregnancy, B) middle pregnancy, C) late pregnancy, and D) averaged over pregnancy and standard deviation scores (SDS) of fetal growth parameters (head circumference, length, and weight) measured during pregnancy by ultrasound and by clinical examination at delivery in the Generation R Study population ($n = 784$). Model adjusted for fetal sex, maternal age (continuous), prepregnancy weight (continuous), height (continuous), education level (categorical), maternal ethnicity (categorical), parity (categorical), smoking (categorical), alcohol use (categorical), folic acid use (categorical), and gestational age at ultrasound or delivery (continuous). This figure includes imputed data. Note: CI, confidence interval; DAPs, dialkyl phosphates; SDS, standard deviation scores.

Table 5. Adjusted difference in fetal head circumference, length, or weight standard deviation score (SDS) at selected weeks gestation in association with pregnancy average urinary organophosphate pesticide metabolite concentrations in a subset of the Generation R Study population from models stratified by infant sex ($n = 398$ males, 386 females).

	Difference in SDS score (95% CI)									
	Total DMPs			Total DEPs			Total DAPs			
Anthropometric parameter/gestational age	Male	Female	p_{int}[a]	Male	Female	p_{int}[a]	Male	Female	p_{int}[a]
Head circumference (weeks)									
20	−0.13 (−0.57, 0.31)	−0.20 (−0.67, 0.28)	0.25	0.06 (−0.12, 0.25)	−0.14 (−0.37, 0.09)	0.80	−0.11 (−0.56, 0.35)	−0.18 (−0.67, 0.30)	0.36
30	−0.01 (−0.36, 0.34)	−0.23 (−0.13, 0.59)		0.07 (−0.07, 0.22)	0.04 (−0.13, 0.21)		0.02 (−0.35, 0.39)	0.20 (−0.16, 0.57)	
40	0.10 (−0.43, 0.64)	−0.66 (0.06, 1.25)		0.08 (−0.16, 0.32)	0.22 (−0.04, 0.47)		0.15 (−0.42, 0.71)	0.59 (0.00, 1.18)	
Length (weeks)									
20	−0.79 (−1.22, −0.37)	−0.18 (−0.63, 0.27)	0.02	−0.16 (−0.34, 0.02)	−0.23 (−0.45, 0.00)	0.49	−0.84 (−1.27, −0.40)	−0.27 (−0.74, 0.19)	0.03
30	−0.38 (−0.69, −0.07)	0.14 (−0.20, 0.48)		−0.04 (−0.17, 0.10)	−0.14 (−0.31, 0.02)		−0.40 (−0.72, −0.07)	0.07 (−0.28, 0.42)	
40	0.03 (−0.43, 0.49)	0.46 (−0.01, 0.92)		0.08 (−0.12, 0.29)	−0.06 (−0.29, 0.17)		0.04 (−0.43, 0.51)	0.42 (−0.06, 0.90)	
Weight (weeks)									
20	−0.61 (−1.00, −0.22)	0.07 (−0.32, 0.46)	<0.01	−0.16 (−0.32, 0.00)	−0.18 (−0.38, 0.01)	0.67	−0.65 (−1.06, −0.25)	0.02 (−0.38, 0.43)	0.01
30	−0.37 (−0.67, −0.07)	0.15 (−0.17, 0.46)		−0.06 (−0.18, 0.06)	−0.14 (−0.30, 0.01)		−0.40 (−0.71, −0.09)	0.11 (−0.21, 0.44)	
40	−0.13 (−0.51, 0.24)	0.22 (−0.19, 0.63)		0.04 (−0.11, 0.20)	−0.11 (−0.32, 0.10)		−0.15 (−0.54, 0.24)	0.20 (−0.22, 0.62)	

Note: Models adjusted for age (continuous), prepregnancy weight (continuous), height (continuous), education level (categorical), ethnicity (categorical), parity (categorical), smoking (categorical), alcohol use (categorical), folic acid use (categorical), and gestational age at ultrasound or delivery (continuous). Models contain an interaction term between exposure concentration and gestational age at ultrasound/delivery, a random intercept for each participant, and a random slope for gestational age at ultrasound/delivery. Models of head circumference, length, and weight include measurements taken at delivery (head circumference, birth length, and birth weight, respectively). CI, confidence interval; DAPs, dialkyl phosphates; DEMs, diethyl phosphates; DMPs, dimethyl phosphates; p_{int}, p for interaction. Values in this table include imputed data.
[a]p-Values for the interaction term between exposure concentration and sex.

greatly across these populations. The average urinary DAP concentration in this study population was 312 nmol/g creatinine, whereas levels in most studies range from 10–100 nmol/g creatinine. Levels may be higher in the Generation R cohort participants due to a higher consumption of fruits in this study population or to the application of larger quantities of pesticides in farms in the Netherlands (van den Dries et al. 2018). The only study to note higher levels was the abovementioned study by Huang et al. (2017), which reported geometric mean concentrations of summed DMP and DEP metabolites as high as 569 and 282 nmol/g creatinine, respectively, in samples analyzed at 11 and 26 weeks of gestation and at delivery. It is not apparent that studies with higher exposure levels are more likely to demonstrate an association with differences in birth outcomes at delivery; however, this might be difficult to detect if all members of the given population are more highly exposed.

Associations observed between urinary DAPs and fetal length and weight in mid-pregnancy were more pronounced in males compared with females. Similar sex differences have been observed in associations between this exposure and neurobehavioral deficits, with males demonstrating stronger associations (Horton et al. 2012; Marks et al. 2010). Placental differences by sex, including epigenetic patterns (Martin et al. 2017), could influence the amount of the toxic compound that is transferred to the fetus and partially explain these differences. Alternatively, the differences observed may be due to the fact that the male fetus is more vulnerable to adverse pregnancy outcomes, particularly to perturbations during their rapid growth in early pregnancy (Eriksson et al. 2010; Pedersen 1980).

The findings from the present study may be difficult to interpret clinically because no associations were detected with size at delivery. However, differences in growth in early pregnancy may be crucial for health outcomes later in life. First-trimester growth restriction is associated with faster weight gain and adverse cardiovascular profiles in school-age children (Jaddoe et al. 2014; Mook-Kanamori et al. 2010). This may be a particularly sensitive time in development, and the consequences of the associations we observed should be investigated in future work.

Our ability to detect differences in growth in association with exposure early in pregnancy could be due to methodological issues as well. Estimated fetal weight as calculated by a combination of ultrasound measures is subject to much more measurement error compared with birth weight (Dudley 2005). However, we would predict that this would lead to improved ability to detect associations between exposure and weight at delivery rather than early in pregnancy. Alternatively, because fetal weight gain occurs primarily in the third trimester, the influence of any error in the estimate of gestational age might be more pronounced toward the end of pregnancy. This could partially explain this difference in our findings based on timing of outcome measurement.

The primary limitation of this study is the nonspecificity of DAPs. Because OPs are metabolized rapidly, these biomarkers remain the best and most commonly used indices of total individual-level exposure (Bravo et al. 2002). However, DAPs reflect human exposure to both the toxic compounds as well as their nontoxic metabolites, which are formed outside the body and can enter the human body through the same exposure routes as the parent compounds. Urinary concentrations of DAPs, therefore, may overestimate exposure to the toxic compounds of interest. This may mean, however, that the associations observed with DAPs in this and other studies may be lower than those that would have been observed with a better estimator of exposure (Sudakin and Stone 2011).

In general, our estimates of exposure are superior to those from other human studies because we measured spot urine

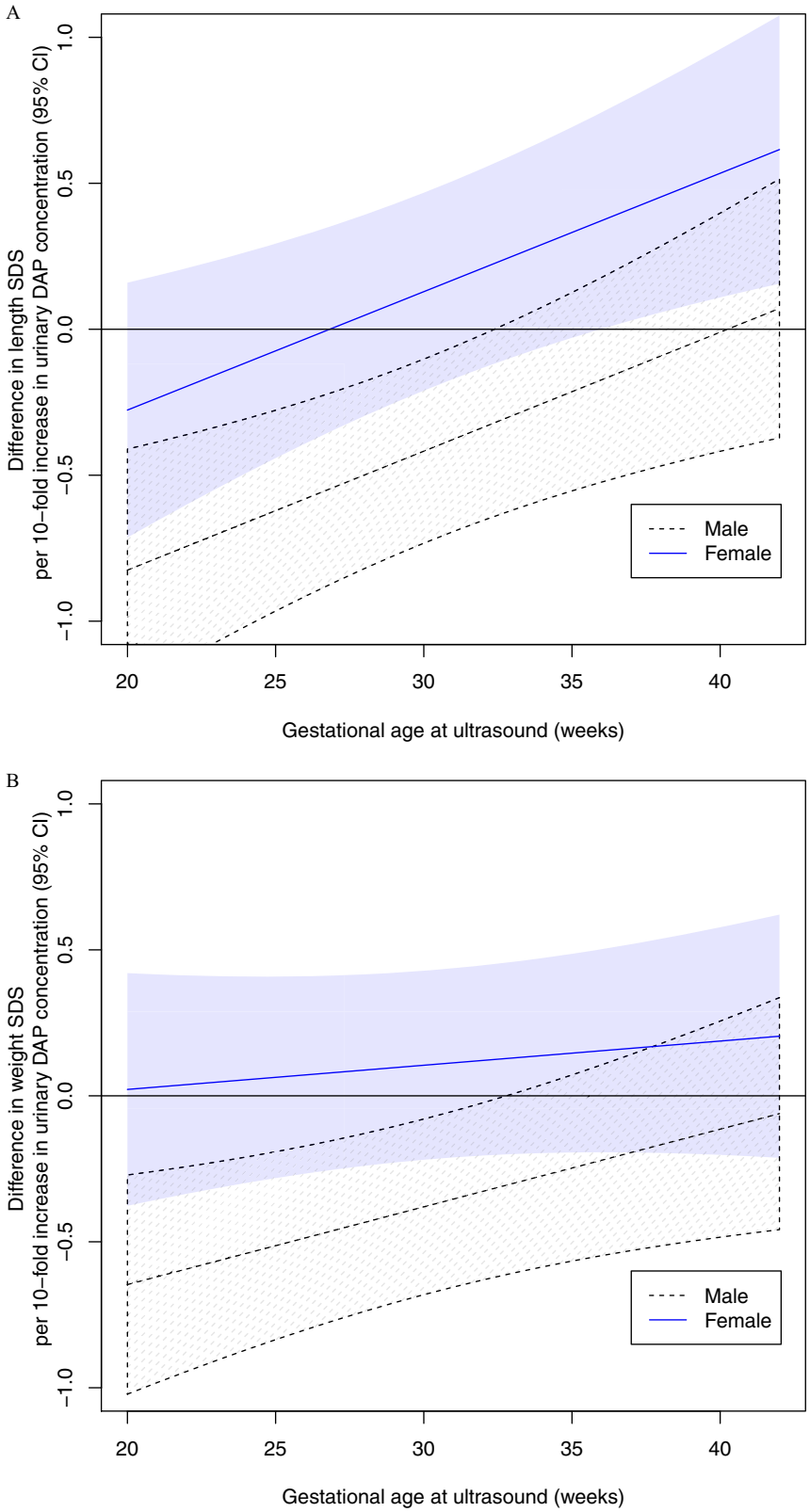

Figure 3. Adjusted and sex-stratified repeated measures associations between pregnancy average total dialkyl phosphate (DAP) concentrations and standard deviation scores of (A) head circumference, (B) length, and (C) weight by gestational age at growth measurement in the Generation R Study population ($n = 784$, 398 males 386 females). Model adjusted for maternal age (continuous), prepregnancy weight (continuous), height (continuous), education level (categorical), maternal ethnicity (categorical), parity (categorical), smoking (categorical), alcohol use (categorical), folic acid use (categorical), and gestational age at ultrasound or delivery (continuous). Models contained an interaction term between exposure concentration and gestational age at ultrasound/delivery, a random intercept for each participant, and a random slope for gestational age at ultrasound/delivery. Main effect and interaction terms (95% CIs) for each plot are as follows: (A) male: −0.35 (−1.35, 0.66); 0.01 (−0.02, 0.05); female: −0.92 (−2.06, 0.22); 0.04 (0.00, 0.08); (B) male: −1.67 (−2.62, −0.72); 0.04 (0.01, 0.08); female: −0.93 (−1.90, 0.04); 0.03 (0.00, 0.07); (C) male: −1.13 (−1.93, −0.33); 0.03 (0.00, 0.05); female: −0.15 (−0.93, 0.64); 0.01 (−0.02, 0.03). This figure includes imputed data. Note: CI, confidence interval; DAPs, dialkyl phosphates; SDS, standard deviation scores.

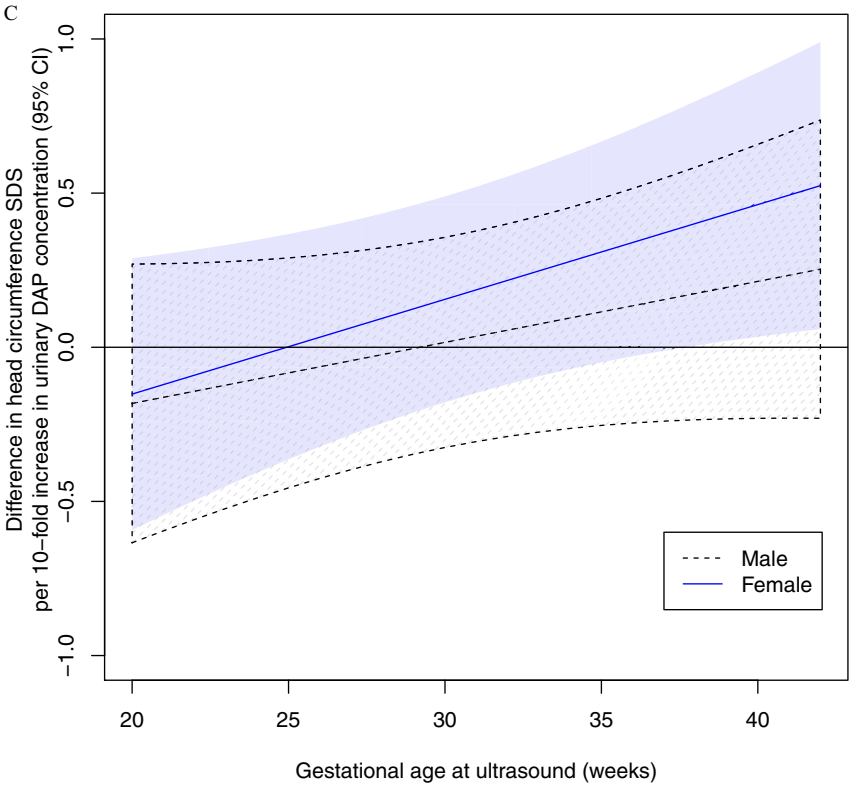

Figure 3. (*Continued.*)

concentrations of DAPs at three time points in pregnancy, creating subject-specific averages that may be a more stable reflection of exposure over time (Spaan et al. 2015). Despite this improvement, the measurement error may have biased our effect estimates toward the null. Indeed, adjusting for measurement error with regression calibration resulted in effect estimates that were farther from the null but more imprecise, illustrating the known trade-off between bias and variance (Carroll et al. 2006).

Our study was also limited by smaller sample sizes with available information on head circumference and body length on neonates. However, we handled missingness by imputing using the MICE procedure, which we previously showed was a suitable approach for fetal growth data (Ferguson et al. 2018). In addition, associations in an unimputed data set were very similar to those shown in our primary results, with the exception of the associations between DAPs and head circumference, which were closer to the null. Last, combining measures of fetal size with measures at delivery in repeated measures models may be problematic because they are measured differently and hence reflect different outcomes. This could be another explanation for the differences we observed between associations with ultrasound measurements in mid-pregnancy vs. anthropometric measurements at delivery. However, we also observed associations closer to the null at late pregnancy with measurements also taken by ultrasound, so we believe this is unlikely to be the case.

The major strengths of our study were the large sample size, the availability of three urinary measurements of DAP metabolites to assess exposure, and the use of repeated ultrasound scans that captured fetal size at multiple time points in pregnancy and in different parameters (e.g., length in addition to weight). This allowed us to investigate associations with OP exposure during gestation that have not been previously examined and enabled detection of decreased fetal growth in early pregnancy in association with exposure.

In summary, urinary biomarkers of OPs were inversely associated with length and weight in mid-pregnancy, with stronger associations observed for exposure biomarkers measured in urine samples collected during early and mid-pregnancy as well as with stronger associations observed in males compared with females. Future research should be directed toward improving the understanding of the consequence of these differences observed on health outcomes later in life.

Acknowledgments

The authors thank the participating parents and their children, general practitioners, hospitals, midwives, and pharmacies for their contribution. Further, the authors also thank F. Pierik for laying the groundwork for the National Institute of Environmental Health Sciences (NIEHS)-TNO-Erasmus University collaboration. This research received financial support from the Intramural Research Program of the NIEHS, National Institutes of Health. The Generation R Study is conducted by the Erasmus Medical Center, Rotterdam, in close collaboration with the Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service, Rotterdam Homecare Foundation, and Stichting Trombosedienst en Artsenlaboratorium Rijnmond.

References

Aguilera I, Garcia-Esteban R, Iñiguez C, Nieuwenhuijsen MJ, Rodríguez À, Paez M, et al. 2010. Prenatal exposure to traffic-related air pollution and ultrasound measures of fetal growth in the INMA Sabadell cohort. Environ Health Perspect 118(5):705–711, PMID: 20103496, https://doi.org/10.1289/ehp.0901228.

Barker DJ. 2006. Adult consequences of fetal growth restriction. Clin Obstet Gynecol 49(2):270–283, PMID: 16721106, https://doi.org/10.1097/00003081-200606000-00009.

Bravo R, Driskell WJ, Whitehead RD Jr, Needham LL, Barr DB. 2002. Quantitation of dialkyl phosphate metabolites of organophosphate pesticides in human urine using GC-MS-MS with isotopic internal standards. J Anal Toxicol 26(5):245–252, PMID: 12166810, https://doi.org/10.1093/jat/26.5.245.

Breslin WJ, Liberacki AB, Dittenber DA, Quast JF. 1996. Evaluation of the developmental and reproductive toxicity of chlorpyrifos in the rat. Fundam Appl Toxicol 29(1):119–130, PMID: 8838647, https://doi.org/10.1006/faat.1996.0013.

Campos É, Freire C. 2016. Exposure to non-persistent pesticides and thyroid function: a systematic review of epidemiological evidence. Int J Hyg Environ Health 219(6):481–497, PMID: 27262599, https://doi.org/10.1016/j.ijheh.2016.05.006.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, FL: Chapman and Hall/CRC.

Casas M, Basagaña X, Sakhi AK, Haug LS, Philippat C, Granum B, et al. 2018. Variability of urinary concentrations of non-persistent chemicals in pregnant women and school-aged children. Environ Int 121(Pt 1):561–573, PMID: 30300814, https://doi.org/10.1016/j.envint.2018.09.046.

Chanda SM, Pope CN. 1996. Neurochemical and neurobehavioral effects of repeated gestational exposure to chlorpyrifos in maternal and developing rats. Pharmacol Biochem Behav 53(4):771–776, PMID: 8801577, https://doi.org/10.1016/0091-3057(95)02105-1.

Dalsager L, Christensen LE, Kongsholm MG, Kyhl HB, Nielsen F, Schoeters G, et al. 2018. Associations of maternal exposure to organophosphate and pyrethroid insecticides and the herbicide 2,4-D with birth outcomes and anogenital distance at 3 months in the Odense Child Cohort. Reprod Toxicol 76:53–62, PMID: 29274948, https://doi.org/10.1016/j.reprotox.2017.12.008.

Dudley NJ. 2005. A systematic review of the ultrasound estimation of fetal weight. Ultrasound Obstet Gynecol 25(1):80–89, PMID: 15505877, https://doi.org/10.1002/uog.1751.

Eriksson JG, Kajantie E, Osmond C, Thornburg K, Barker DJ. 2010. Boys live dangerously in the womb. Am J Hum Biol 22(3):330–335, PMID: 19844898, https://doi.org/10.1002/ajhb.20995.

Eskenazi B, Bradman A, Castorina R. 1999. Exposures of children to organophosphate pesticides and their potential adverse health effects. Environ Health Perspect 107(Suppl 3):409–419, PMID: 10346990, https://doi.org/10.1289/ehp.99107s3409.

Eskenazi B, Harley K, Bradman A, Weltzien E, Jewell NP, Barr DB, et al. 2004. Association of *in utero* organophosphate pesticide exposure and fetal growth and length of gestation in an agricultural population. Environ Health Perspect 112(10):1116–1124, PMID: 15238287, https://doi.org/10.1289/ehp.6789.

Ferguson KK, Meeker JD, Cantonwine DE, Chen Y-H, Mukherjee B, McElrath TF. 2016. Urinary phthalate metabolite and bisphenol A associations with ultrasound and delivery indices of fetal growth. Environ Int 94:531–537, PMID: 27320326, https://doi.org/10.1016/j.envint.2016.06.013.

Ferguson KK, Yu Y, Cantonwine DE, McElrath TF, Meeker JD, Mukherjee B. 2018. Foetal ultrasound measurement imputations based on growth curves versus multiple imputation chained equation (MICE). Paediatr Perinat Epidemiol 32(5):469–473, PMID: 30016545, https://doi.org/10.1111/ppe.12486.

Gaillard R, Steegers EA, de Jongste JC, Hofman A, Jaddoe VW. 2014. Tracking of fetal growth characteristics during different trimesters and the risks of adverse birth outcomes. Int J Epidemiol 43(4):1140–1153, PMID: 24603318, https://doi.org/10.1093/ije/dyu036.

Gardosi J, Chang A, Kalyan B, Sahota D, Symonds E. 1992. Customised antenatal growth charts. Lancet 339(8788):283–287, PMID: 1346292, https://doi.org/10.1016/0140-6736(92)91342-6.

Gishti O, Gaillard R, Manniesing R, Abrahamse-Berkeveld M, van der Beek EM, Heppe DH, et al. 2014. Fetal and infant growth patterns associated with total and abdominal fat distribution in school-age children. J Clin Endocrinol Metab 99(7):2557–2566, PMID: 24712569, https://doi.org/10.1210/jc.2013-4345.

Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. 1985. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. Am J Obstet Gynecol 151(3):333–337, PMID: 3881966, https://doi.org/10.1016/0002-9378(85)90298-4.

Haines DA, Murray J. 2012. Human biomonitoring of environmental chemicals—early results of the 2007–2009 Canadian Health Measures Survey for males and females. Int J Hyg Environ Health 215(2):133–137, PMID: 22001329, https://doi.org/10.1016/j.ijheh.2011.09.008.

Hardin JW, Schmiediche H, Carroll R. 2003. The regression-calibration method for fitting generalized linear models with additive measurement error. Stata J 3(4):361–372, https://doi.org/10.1177/1536867X0400300407.

Harley KG, Engel SM, Vedar MG, Eskenazi B, Whyatt RM, Lanphear BP, et al. 2016. Prenatal exposure to organophosphorous pesticides and fetal growth: Pooled results from four longitudinal birth cohort studies. Environ Health Perspect 124(7):1084–1092, PMID: 26685281, https://doi.org/10.1289/ehp.1409362.

Henrichs J, Schenk JJ, Barendregt CS, Schmidt HG, Steegers EA, Hofman A, et al. 2010. Fetal growth from mid- to late pregnancy is associated with infant development: the Generation R Study. Dev Med Child Neurol 52(7):644–651, PMID: 19832882, https://doi.org/10.1111/j.1469-8749.2009.03513.x.

Horton MK, Kahn LG, Perera F, Barr DB, Rauh V. 2012. Does the home environment and the sex of the child modify the adverse effects of prenatal exposure to chlorpyrifos on child working memory? Neurotoxicol Teratol 34(5):534–541, PMID: 22824009, https://doi.org/10.1016/j.ntt.2012.07.004.

Huang Y-F, Pan W-C, Tsai Y-A, Chang C-H, Chen P-J, Shao Y-S, et al. 2017. Concurrent exposures to nonylphenol, bisphenol A, phthalates, and organophosphate pesticides on birth outcomes: a cohort study in Taipei, Taiwan. Sci Total Environ 607–608:1126–1135, PMID: 28724251, https://doi.org/10.1016/j.scitotenv.2017.07.092.

Jaddoe VW, de Jonge LL, Hofman A, Franco OH, Steegers EA, Gaillard R. 2014. First trimester fetal growth restriction and cardiovascular risk factors in school age children: population based cohort study. BMJ 348:g14, PMID: 24458585, https://doi.org/10.1136/bmj.g14.

Jusko TA, van den Dries MA, Pronk A, Shaw PA, Guxens M, Spaan S, et al. 2019. Organophosphate pesticide metabolite concentrations in urine during pregnancy and offspring nonverbal IQ at age 6 years. Environ Health Perspect 127(1):017007, PMID: 30688513, https://doi.org/10.1289/EHP3024.

Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van IJzendoorn MH, et al. 2016. The Generation R Study: design and cohort update 2017. Eur J Epidemiol 31(12):1243–1264, PMID: 28070760, https://doi.org/10.1007/s10654-016-0224-9.

Kruithof CJ, Kooijman MN, van Duijn CM, Franco OH, de Jongste JC, Klaver CC, et al. 2014. The Generation R Study: biobank update 2015. Eur J Epidemiol 29(12):911–927, PMID: 25527369, https://doi.org/10.1007/s10654-014-9980-6.

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. 2010. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34(8):816–834, PMID: 21058334, https://doi.org/10.1002/gepi.20533.

Liu P, Wu C, Chang X, Qi X, Zheng M, Zhou Z. 2016. Adverse associations of both prenatal and postnatal exposure to organophosphorous pesticides with infant neurodevelopment in an agricultural area of Jiangsu Province, China. Environ Health Perspect 124(10):1637–1643, PMID: 27153333, https://doi.org/10.1289/EHP196.

Llop S, Murcia M, Iñiguez C, Roca M, González L, Yusà V, et al. 2017. Distributions and determinants of urinary biomarkers of organophosphate pesticide exposure in a prospective Spanish birth cohort study. Environ Health 16:(1):46, PMID: 28514952, https://doi.org/10.1186/s12940-017-0255-z.

Lu C, Barr DB, Pearson MA, Waller L. 2008. Dietary intake and its contribution to longitudinal organophosphorus pesticide exposure in urban/suburban children. Environ Health Perspect 116(4):537–542, PMID: 18414640, https://doi.org/10.1289/ehp.10912.

Marks AR, Harley K, Bradman A, Kogut K, Barr DB, Johnson C, et al. 2010. Organophosphate pesticide exposure and attention in young Mexican-American children: the CHAMACOS study. Environ Health Perspect 118(12):1768–1774, PMID: 21126939, https://doi.org/10.1289/ehp.1002056.

Martin E, Smeester L, Bommarito PA, Grace MR, Boggess K, Kuban K, et al. 2017. Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. Epigenomics 9(3):267–278, PMID: 28234023, https://doi.org/10.2217/epi-2016-0132.

Maurissen JP, Hoberman AM, Garman RH, Hanley TR Jr. 2000. Lack of selective developmental neurotoxicity in rat pups from dams treated by gavage with chlorpyrifos. Toxicol Sci 57(2):250–263, PMID: 11006355, https://doi.org/10.1093/toxsci/57.2.250.

Miller SL, Huppi PS, Mallard C. 2016. The consequences of fetal growth restriction on brain structure and neurodevelopmental outcome. J Physiol 594(4):807–823, PMID: 26607046, https://doi.org/10.1113/JP271402.

Mook-Kanamori DO, Steegers EA, Eilers PH, Raat H, Hofman A, Jaddoe VW. 2010. Risk factors and outcomes associated with first-trimester fetal growth restriction. JAMA 303(6):527–534, PMID: 20145229, https://doi.org/10.1001/jama.2010.78.

Muto M, Lobelle JF Jr, Bidanset J, Wurpel J. 1992. Embryotoxicity and neurotoxicity in rats associated with prenatal exposure to DURSBAN. Vet Hum Toxicol 34(6):498–501, PMID: 1283795.

Naksen W, Prapamontol T, Mangklabruks A, Chantara S, Thavornyutikarn P, Srinual N, et al. 2015. Associations of maternal organophosphate pesticide exposure and PON1 activity with birth outcomes in SAWASDEE birth cohort, Thailand. Environ Res 142:288–296, PMID: 26186137, https://doi.org/10.1016/j.envres.2015.06.035.

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 2015. A global reference for human genetic variation. Nature 526(7571):68–74, PMID: 26432245, https://doi.org/10.1038/nature15393.

Pedersen JF. 1980. Ultrasound evidence of sexual difference in fetal size in first trimester. Br Med J 281(6250):1253, PMID: 7427655, https://doi.org/10.1136/bmj.281.6250.1253.

Philippat C, Botton J, Calafat AM, Ye X, Charles M-A, Slama R. 2014. Prenatal exposure to phenols and growth in boys. Epidemiology 25(5):625–635, PMID: 25061923, https://doi.org/10.1097/EDE.0000000000000132.

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2014. R core team (2014) nlme: Linear and nonlinear mixed effects models. R package version 3.1-117. https://CRAN.R-project.org/package=nlme [accessed 8 August 2019].

Rauch SA, Braun JM, Barr DB, Calafat AM, Khoury J, Montesano MA, et al. 2012. Associations of prenatal exposure to organophosphate pesticide metabolites

with gestational age and birth weight. Environ Health Perspect 120(7):1055–1060, PMID: 22476135, https://doi.org/10.1289/ehp.1104615.

Resnik R. 2018. Fetal growth restriction: Evaluation and management. https://www.uptodate.com/contents/fetal-growth-restriction-evaluation-and-management [accessed 8 August 2019].

Snijder CA, Heederik D, Pierik FH, Hofman A, Jaddoe VW, Koch HM, et al. 2013. Fetal growth and prenatal exposure to bisphenol A: the Generation R Study. Environ Health Perspect 121(3):393–398, PMID: 23459363, https://doi.org/10.1289/ehp.1205296.

Sokoloff K, Fraser W, Arbuckle TE, Fisher M, Gaudreau E, LeBlanc A, et al. 2016. Determinants of urinary concentrations of dialkyl phosphates among pregnant women in Canada—results from the MIREC study. Environ Int 94:133–140, PMID: 27243443, https://doi.org/10.1016/j.envint.2016.05.015.

Song X, Seidler FJ, Saleh JL, Zhang J, Padilla S, Slotkin TA. 1997. Cellular mechanisms for developmental toxicity of chlorpyrifos: targeting the adenylyl cyclase signaling cascade. Toxicol Appl Pharmacol 145(1):158–174, PMID: 9221834, https://doi.org/10.1006/taap.1997.8171.

Spaan S, Pronk A, Koch HM, Jusko TA, Jaddoe VW, Shaw PA, et al. 2015. Reliability of concentrations of organophosphate pesticide metabolites in serial urine specimens from pregnancy in the Generation R Study. J Expo Sci Environ Epidemiol 25(3):286–294, PMID: 25515376, https://doi.org/10.1038/jes.2014.81.

Strand LB, Barnett AG, Tong S. 2011. The influence of season and ambient temperature on birth outcomes: a review of the epidemiological literature. Environ. Res 111(3):451–462, PMID: 21333980, https://doi.org/10.1016/j.envres.2011.01.023.

Sudakin DL, Stone DL. 2011. Dialkyl phosphates as biomarkers of organophosphates: the current divide between epidemiology and clinical toxicology. Clin Toxicol (Phila) 49(9):771–781, PMID: 22077242, https://doi.org/10.3109/15563650.2011.624101.

Timmermans S, Jaddoe VW, Hofman A, Steegers-Theunissen RP, Steegers EA. 2009. Periconception folic acid supplementation, fetal growth and the risks of low birth weight and preterm birth: The Generation R Study. Br J Nutr 102(5):777–785, PMID: 19327193, https://doi.org/10.1017/S0007114509288994.

Timmermans S, Jaddoe VW, Mackenbach JP, Hofman A, Steegers-Theunissen RP, Steegers EA. 2008. Determinants of folic acid use in early pregnancy in a multi-ethnic urban population in the Netherlands: the Generation R Study. Preventive Medicine 47(4):427–432, PMID: 18644404, https://doi.org/10.1016/j.ypmed.2008.06.014.

van Buuren S, Groothuis-Oudshoorn K. 2010. mice: multivariate imputation by chained equations in R. J Stat Soft 45(3):1–67, https://doi.org/10.18637/jss.v045.i03.

van den Dries MA, Pronk A, Guxens M, Spaan S, Voortman T, Jaddoe VW, et al. 2018. Determinants of organophosphate pesticide exposure in pregnant women: a population-based cohort study in the Netherlands. Int J Hyg Environ Health 221(3):489–501, PMID: 29499913, https://doi.org/10.1016/j.ijheh.2018.01.013.

Wang P, Tian Y, Wang X-J, Gao Y, Shi R, Wang G-Q, et al. 2012. Organophosphate pesticide exposure and perinatal outcomes in Shanghai, China. Environ Int 42:100–104, PMID: 21601922, https://doi.org/10.1016/j.envint.2011.04.015.

Wolff MS, Engel S, Berkowitz G, Teitelbaum S, Siskind J, Barr DB, et al. 2007. Prenatal pesticide and PCB exposures and birth outcomes. Pediatr Res 61(2):243–250, PMID: 17237730, https://doi.org/10.1203/pdr.0b013e31802d77f0.

Woods MM, Lanphear BP, Braun JM, McCandless LC. 2017. Gestational exposure to endocrine disrupting chemicals in relation to infant birth weight: a Bayesian analysis of the HOME Study. Environ Health 16(1):115, PMID: 29078782, https://doi.org/10.1186/s12940-017-0332-3.