Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
Article 1: Whyatt et al. (2004) Columbia U. (N=314)	Birth length, Birth weight, head circumference	Four cord plasma chlorpyrifos exposure groups and four chlorpyrifos and diazinon exposure groups. Chlorpyrifos only categories, Group 1: levels below LOD (32% of participants); Group 2: lowest 1/3 of detectable levels (20 %); Group 3: middle 1/3 (24%), Group 4: highest 1/3 (25%.) Chlorpyrifos and diazinon together: Group 1: 26%, Group 2: 22 %, Group 3: 26%, Group 4: 26%.	Gestational age, maternal pre-pregnancy weight, maternal net pregnancy weight gain, gender of newborn, parity, race/ethnicity, ETS in home, season, cesarean section	For each log unit increase in cord plasma chlorpyrifos levels, birth weight decreased by 42.6 g (95% CI: –81.8 to –3.8) and birth length decreased by 0.24 cm (95% CI: –0.47 to –0.01). Birth weight averaged 186.3 g less (95% CI: –375.2 to –45.5) among newborns with the highest compared with lowest 26% of exposure levels ($p = 0.01$).	Associations between birth weight and length and cord plasma chlorpyrifos were statistically significant ($p \leq 0.007$) among newborns born before the January 2001 policy change. Among newborns born after January 2001, exposure levels were substantially lower, and no associations with fetal growth outcomes were observed ($p > 0.8$).	Strengths: prospective nature of the study; direct measurement of chlorpyrifos in cord blood and personal air samples, rather than non-specific markers of organophosphate pesticide exposure; consideration of other pesticides and environmental contaminants as covariates in the multivariate models. Limitations: single exposure sampling period; the authors did not present nor discuss regression diagnostics to assess the degree to which their models met or violated the assumptions implicit in linear models.
Article 2: Berkowitz et al.(2004) Mt. Sinai (N=404)	Birth length, birth weight, head circumference, gestational age	LOD: 11 ug/L (57% <LOD TCPy)	Race/ethnicity, infant sex, and gestational age. The authors also controlled for birth weight or birth length in their assessment of head circumference and pesticide exposure.	Mean levels of birth weight, length, head circumference, and gestational age did not differ between those with urinary pesticide metabolite levels below and above the level of detection. Similarly, no statistically significant associations were observed between reported pesticide exposure and mean indices of fetal growth and gestational age.	PON1 activity also predictor of smaller head circumference; creatinine corrected	Very well conducted study with numerous strengths and very few weaknesses. The questionnaire-based pesticide exposure questions are subject to imperfect recall. Errors would, on average, attenuate associations between these exposure metrics and fetal development. Recall-based exposure assessments were fortified by objective measures of pesticides/pesticide

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
						metabolites. A metabolite specific for chlorpyrifos (TCPy) was assessed. Statistical analysis was appropriate. Observed mean reductions in the outcome parameters appear to be small in magnitude and may be of little clinical significance. Assessment and control for confounding were appropriate. However, confounding by unmeasured (and mismeasured) risk factors for abnormal growth that are related to pesticide exposure would bias the results in this study. Such factors may be related to socioeconomic status of the study participants, a cofactor which is difficult to define, no less measure in an epidemiologic study. Limited external validity (generalizability) due to the particular study population recruited and the numerous exclusion criteria applied.
Article 3: Eskenazi et al. (2004) CHAMACOS (N=488)	Birth length, birth weight, head circumference, Gestation Length, Ponderal index	Total DAPs: median 136 nmol/L (range: 10–6,854); DEP: median 22 nmol/L (range: 2–680 nmol/L); TCPy: median 3.3 nmol/L (range: 0.2–56.1nmol/L) (76% >LOD)	Gestational age, gestational age squared, maternal age, pregnancy weight gain, week of initiating prenatal care, parity, infant sex, mother's country of birth, body mass index, family income, poverty level, smoking, alcohol, illicit drug use, environmental tobacco smoke,	Decreases in gestational duration associated with two measures of in utero pesticide exposure: levels of metabolites of dimethyl phosphate pesticide compounds and whole blood ChE.	Maternal urine collection averaged wks.14, 26, not creatinine-corrected	Strengths in the study design include the longitudinal design, the use of multiple exposure biomarkers, including quantification of non-specific (DAPs), chlorpyrifos-specific (TCPy) metabolites, and other environmental co-exposures. A reasonable set of exclusion

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
			caffeine, history of low birth weight, and history of pre-term delivery.			criteria were applied.The selection of the CHAMACOS population, which consists mostly of children from low-income families, served to increase the relative statistical efficiency of the study, as this population is at high risk of neurodevelopmental deficits, compared to the general population.The statistical analysis used to assess the associations between the markers of exposure and neurodevelopment were appropriate.Errors in the assignment of exposure in this prospective study will likely have resulted in attenuation of observed associations.
Article 4: Harley et al. (2011) CHAMACOS (N=329)	Birth length, birth weight, head circumference, gestational age	The geometric mean for the DAP concentrations during pregnancy (for the average of the two sampling periods) was 146 nmol/L (95% CI: 133, 160); of this, a larger proportion was DMP metabolites (GM = 109 nmol/L; 95% CI =98, 120) than DEPs (GM= 23 nmol/L; 95% CI = 21, 25). Allele frequencies: PON1 192 Q allele= 50%; PON1 -108 T allele= 46%.Mean arylesterase activity: For infants: 33.6 U/mL (SD = 16) For mothers: 136.6 U/mL (SD = 44). Mean paraoxonase	Maternal intelligence (Peabody Picture Vocabulary Test (PPVT)), measures of how stimulating the environment is, and known or suspected neurotoxins were measured prenatally. To measure the quality and extent of stimulation available to a child in the home environment, the Infant-Toddler HOME (Home Observation for Measurement of the Environment) inventory was completed at the 6-month, 1, 2, 3.5, 5, and 7 year visits; known or suspected neurotoxicants, polybrominated diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs), p,p'-dichlorodiphenyltrichlorethane (DDT), p,p'-dichlorodiphenyltrichlorethylene (DDE), and lead.	The authors observed evidence of an association between prenatal exposure to OP pesticides as measured by urinary DAP metabolites in women during pregnancy, is associated with decreased cognitive functioning in children at age 7.	Infants whose PON1 genotype and enzyme activity levels suggested that they might be more susceptible tothe effects of OP pesticide exposure had decreased fetal growth and length of gestation. PON1 may be a contributing factor to preterm or low birth weight birth.	This study has many strengths, the longitudinal design, the measurement of urinary DAP at multiple time points and following children to age seven when tests of cognitive function are reportedly more reliable. The authors were able to adjust for or consider many factors related to cognitive function, such as prenatal exposure to other environmental agents, socioeconomic indicators, maternal intelligence and education, and child stimulation. The cohort had a relatively homogenous socioeconomic profile, reducing the potential for uncontrolled confounding.

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
		activity: For infants: 256.6 U/L (SD = 165); For mothers: 989.0 U/L (SD = 616).				
Article 5: Engel et al. (2007) Mt. Sinai (N=311)	Brazelton Neonatal Behavioral Assessment Scale (BNBAS) , primitive reflexes (neurological integrity) measured before hospital discharge.	Diethylphosphates (DEP): 24.7 nmol/L; Total DAP: 82 nmol/L	Maternal age, race, marital status, education, cesarean delivery, delivery anesthesia, infant age at examination, infant gender, infant jaundice, smoking (yes/no), alcohol consumption, caffeine consumption, illicit drug use during pregnancy, and the examiner.	No adverse associations were found for DAPs and any measured behavior. Relative to the first quartile, quartiles 2–4 of total DEPs, DMPs, and DAPs were associated with an increased proportion of abnormal reflexes, although the associations did not increase monotonically and varied in their strength and precision.	Used non-specific biomarker DEP/DAP	This was a well conducted prospective study conducted in a young, predominantly minority population. The study design, analytic approach, and statistical analyses were appropriate. Pesticide metabolites evaluated are not specific for chlorpyrifos. The BNBAS was administered before hospital discharge only on a subset of children in the cohort (n =311/404). Factors related to weekend delivery (e.g., fewer inductions) would be underrepresented among the tested subjects, and may induce bias, reduce the degree of precision with which associations were estimated, and limit the generalizability of the study findings.The statistical analysis was largely appropriate. Imputing of missing data may affect precision of association estimates and result in attenuated effect estimates as a result of exposure measurement error.Assessment and control for confounding were appropriate. However, confounding by unmeasured (and mi-asured) risk factors for abnormal growth that are related to pesticide exposure would bias the results in this study. Such factors may be related to socioeconomic status of the study participants, a

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
						cofactor which is difficult to define, no less measure in an epidemiologic study.
Article 6: Young et al. (2005) CHAMACOS (N=381)	Neurodevelopment, Brazelton Neonatal Behavioral AssessmentScale (BNBAS), abnormal reflexes	DAP (ave. during pregnancy): median 222nmol/L (range: 7–21,867 nmol/L); DEP (ave. pregnancy): median 21nmol/L (range: 2–680 nmol/L)	Maternal age, BMI, any smoking/alcohol/drug use during pregnancy, gestational age at which prenatal care was initiated, total number of prenatal care visits, mean pregnancy blood pressure, parity, method of delivery, general anesthesia used during delivery, breastfeeding initiated after delivery, poverty level, infant sex, age in days at BNBAS, minutes since last feed at BNBAS, and BNBAS examiner.	Among the >3 day old infants, increasing average prenatal urinary metabolite levels were associated with both an increase in number of abnormal reflexes (total DAP: adjusted beta = 0.53, 95% CI = 0.23, 0.82; dimethyls: adjusted beta = 0.41, 95% CI = 0.12, 0.69; diethyls: adjusted beta = 0.37, 95% CI = 0.09, 0.64), and the proportion of infants with more than three abnormal reflexes (total DAP: adjusted OR = 4.9, 95% CI = 1.5, 16.1; dimethyls: adjusted OR = 3.2, 95% CI = 1.1, 9.8; diethyls: adjusted OR = 3.4, 95% CI = 1.2, 9.9).	Associations seen pre-natal OP, not post-natal OP exposure, Maternal urine collection averaged wks.14, 26	Strengths: Longitudinal design, measurement and consideration of many confounders, the prenatal exposure measures were, with some exceptions, the average of two measurements, and thus may better reflect chronic exposure during the pregnancyWeaknesses: Potential for residual confounding, potential for exposure misclassification, potential selection bias (not all children had the outcome measures), study population not generalizable to the whole US.
Article 7: Rauh et al. (2006) Columbia U. (N=254)	Neurodevelopment: The Bayley Scales of Infant Development II (BSID-II), Mental Development Index (MDI) and Psychomotor Development Index (PDI) at 12, 24, and 36 months of age. • Behavior: Child Behavior Checklist (CBCL) at 12, 24, and 36 months . • Quality of the child-care environment: The Home Observation for Measurement of the Environment (HOME)	Exposure levels were categorized as low (≤ 6.17 pg/g) or high (>6.17 pg/g)	Data were collected regarding lead exposure, demographics, education and occupational history, income, active and passive smoking, alcohol and drug use during pregnancy, and residential pesticide use. Final models included prenatal environmental tobacco smoke (ETS) exposure, gender, ethnicity, gestational age at birth, quality of home care-taking environment, maternal education, and maternal IQ.	At the 36 month milestone, the likelihood of highly exposed children developing mental delays were 2.4 times greater (95% CI: 1.12-5.08, p = .02) and motor delays were 4.9 greater (95% CI: 1.78-13.72; p = .002) than those with lower prenatal exposure. The GLM analysis for PDI scores showed a significant effect of chlorpyrifos exposure over time with an estimated deficit of approximately 7 points by age 36 months (p = .01).	The authors summarize three main findings: 1) by age 3, the more highly exposed children demonstrated mental and motor delays; 2) the observed developmental trajectories for PDI and MDI scores confirmed that the adverse impact on cognitive and motor development increased over time; and 3) by age 3, highly exposed children were more likely to demonstrate clinically significant attention problems.	• Only 53% of the children reached the three year milestone with study data collected. It is unclear what percentage of these children did not survive, were lost to follow-up, or too sick to participate. • Reliance on a single exposure level (prenatal/cord blood.) • No control for exposure over the subsequent 3 years • Creation of a dichotomous exposure variable brings limitations due to the amount of within-group variation. • Limitations of the sensitivity and predictive validity of the developmental tests, especially among children less than 3 years of age. • No discussion of whether this 7-point deficit is

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
						clinically relevant. • Due to the pervasive, non-specific nature of neurological effects, it is difficult to attribute causality.
Article 8: Lovasi et al. 2010 Columbai U. (N=266)	Bayley scores (MDI/PDI) 12 mo., 24 mo., 36 mo.	N/A	Neighborhood characteristics: The percentage of housing units without complete plumbing, the percentage of vacant housing units, the percentage of residents below the federal poverty line, the percentage of residents older than 25 years of age who completed high school, the percentage of households receiving public assistance, the percentage of housing units with one or more residents per room, racial composition, the percentage of residents born outside the United States, the percentage of Spanish-speaking residents, and the percentage of residents who were linguistically isolated	Neighborhood characteristics did not confound the observed association between chlorpyrifos levels and cognitive development.	Hierarchal regression analysis of potential confounding by SES	Direct measurement of chlorpyrifos.The statistical analyses were generally appropriate. Missing data on covariates were estimated using multiple imputation, and the variance estimates presented appropriately reflect the degree of uncertainty caused by missing covariate data. Robust standard errors were used.The setting of the investigation in a sample drawn from low-income African American and Dominican communities is both a strength (increases the power, restriction of confounders) and a limitation of the study (reduced generalizability).
Article 9: Engel et al. (2011) Mt. Sinai (N=276)	Bayley scores (MDI/PDI) at 12 mo., 24 mo.	DEP: 24.7 nmol/L; Total DAP: 82 nmol/L (same as Engel 2007)	Maternal age, race/ethnicity, marital status, education, breast-feeding, child sex, alcohol, smoking, or drug use during pregnancy, maternal IQ, a score based on assessment of the home environment (HOME), season of urine collection, language spoken in the home, age at testing, examiner, and urinary creatinine level.	An observed association between prenatal total dialkylphosphate metabolite level and a decrement in mental development at 12 months among blacks and Hispanics.	Used non-specific biomarker DEP/DAP; some evidence of effect modification by PON1 genotype	Limitations include use of non-specific markers of chlorpyrifos pesticide exposure (DAPs), use of only a single (third-trimester) urine sample, and the large proportion of loss to follow-up. Statistical analysis was appropriate. Imputing of missing data may affect precision of association estimates and result in attenuated effect estimates as a result of exposure measurement error, although these are offset by the further categorization of the exposure levels (at the median). However, binning of exposure levels reduces

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
						precision, relative to a continuously distributed measure of exposure. Assessment and control for confounding were appropriate. However, confounding by unmeasured (and mismeasured) risk factors for abnormal growth that are related to pesticide exposure would bias the results in this study. Such factors may be related to socioeconomic status of the study participants, a cofactor which is difficult to define, no less measure in an epidemiologic study.
Article 10: Eskenazi et al. (2007) CHAMACOS (N=372)	Neurodevelopment, Bayley Index (MDI, PDI), Maternal behavior checklist at 6, 12, and 24 months	DEP: geom. Mean in mother 18.1 nmol/L (95% CI 16.7–19.7); DEP geom mean in child at 24 months 10.5 nmol/L (95% CI 8.8–12.6); TCPy median 3.54 ug/l	Psychometrician, location of assessment, exact age at assessment, sex, breast-feeding duration (months), HOME score, and household income, parity, maternal PPVT, maternal age, education, depressive symptoms, active/passive smoking exposure during pregnancy, regular alcohol use during pregnancy, marital status, father's presence in home, housing density, maternal work status, ≥ 15 hours out-of-home childcare/week, birth weight, gestational age, abnormal reflexes, PCBs, lead, DDT, β-hexachlorocyclohexane, and hexa-chlorobenzene	DAP metabolite levels during pregnancy, particularly from dimethyl phosphate pesticides, may be negatively associated at 24 months with mental development (MDI) on the Bayley Scales and an increase in risk of maternally reported PDD.	No strong associations identified with DE or TCPy, Maternal urine collection averaged wks.14, 26	Strengths: Longitudinal design, measurement and consideration of many confounders (including other environmental chemicals); the prenatal exposure measures were, with some exceptions, the average of two measurements, and thus may better reflect chronic exposure during the pregnancy Weaknesses: Potential for residual confounding, potential for exposure misclassification, potential selection bias (not all children had the outcome measures), study population not generalizable to the whole US.
Article 11: Eskenazi et al. (2010) CHAMACOS (N=371)	Neurodevelopment, Bayley Index (MDI, PDI), Maternal behavior checklist at 6, 12, and 24 months, PON1 gene and enzyme levels	The geometric mean for the DAP concentrations during pregnancy (for the average of the two sampling periods) was 146 nmol/L (95% CI:	Psychometrician, location of assessment, exact age at assessment, sex, breast-feeding duration (months), HOME score, and household income, parity, maternal PPVT, maternal age, education, depressive	Decrease MDI (24 mo.) $PON1_{108TT}$ –5.7 (–9.0 to –2.5) 9=0.01; Decrease PDI (24 mo.) $PON1_{108TT}$ –2.8 (–5.7 to 0.2) p=0.07; increased odds PDD 2.0 (0.8 to 5.1) p=0.14; no association $PON1_{192}$; no	In this study population, evidence PON1 may influence MDI score, but not PDI or PDD risk at two-years. Non-significant evidence of decreasing MDI score by increasing DAP levels across	Strengths: Longitudinal design, measurement and consideration of many confounders (including other environmental chemicals); the prenatal exposure measures were, with some exceptions, the average of two measurements,

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
		133, 160); of this, a larger proportion was DMP metabolites (GM = 109 nmol/L; 95% CI =98, 120) than DEPs (GM= 23 nmol/L; 95% CI = 21, 25). Allele frequencies: PON1 192 Q allele= 50%; PON1 -108 T allele= 46%.Mean arylesterase activity: For infants: 33.6 U/mL (SD = 16) For mothers: 136.6 U/mL (SD = 44). Mean paraoxonase activity: For infants: 256.6 U/L (SD = 165); For mothers: 989.0 U/L (SD = 616).	symptoms, active/passive smoking exposure during pregnancy, regular alcohol use during pregnancy, marital status, father's presence in home, housing density, maternal work status, ≥ 15 hours out-of-home childcare/week, birth weight, gestational age, abnormal reflexes, PCBs, lead, DDT, β-hexachlorocyclohexane, and hexa-chlorobenzene	association PON1 activity measured newborn, 2 years, maternal and MDI, PDI, PDD. Evidence of decreasing MDI score by number of $PON1_{108}$ variant alleles: $PON1_{108CC}$ –3.2 (–9.8 to 3.5), CT –3.7 (–8.0 to 0.6), TT –5.5 (–11.1 to 0.1), p-interaction 0.98.	strata of the number of $PON1_{108}$ variant alleles, interaction non-significant. Similar trend with prenatal DEP levels and MDI, PDI by PON108 alleles, but less pronounced. Overall, limited, non-definitive evidence of effect modification by PON1 status in the relation between mental and psychomotor effects and prenatal DAPs.	and thus may better reflect chronic exposure during the pregnancyWeaknesses: Potential for residual confounding, potential for exposure misclassification, potential selection bias (not all children had the outcome measures), study population not generalizable to the whole US. Study may be under-powered to evaluate effect modification by *PON1* status.
Article 12: Marks et al. (2010) CHAMACOS (N=348)	CBCL; K-CPT; ADHD confidence index; Hillside behavioral rating scale; composite ADHD indicator	DAP (geom mean) pregnancy 109.0 nmol/L; DEP 17.7 nmol/L	Psychometrician, exact age at assessment, sex, maternal education, depressive symptoms, PPVT (continuous), ≥ 15 hr out-of-home child care/week, breast feeding duration (months), maternal age, parity, marital status, active/passive smoking exposure and regular alcohol use during pregnancy, presence of father in home, maternal work status, and household income	Prenatal DAPs were non-significantly associated with maternal report of attention problems and ADHD at age 3.5 years, but were significantly related at age 5 years [CBCL attention problems: β = 0.7 points; 95% confidence interval (CI), 0.2-1.2; ADHD: β = 1.3; 95% CI, 0.4-2.1].	Marked effect modification by gender: 11-fold increase ADHD composite indicator in boys, less than 2-fold in girls, however unstable estimates; weak evidence of association DAPs at 3.5, 5 yrs. and attention	Strengths: Longitudinal design, measurement and consideration of many confounders (including other environmental chemicals); the prenatal exposure measures were, with some exceptions, the average of two measurements, and thus may better reflect chronic exposure during the pregnancy Weaknesses: Potential for residual confounding, potential for exposure misclassification, potential selection bias (not all children had the outcome measures), study population not generalizable to the whole US.
Article 13: Rauh et al. (2011) Columbia U. (N=265)	• Wechsler Scales of Intelligence for Children (WISC-IV) • Child Behavior	• Chlorpyrifos levels in umbilical cord blood samples, N=256 newborns• If no cord	Data were collected regarding lead exposure, demographics, education and occupational history, income, active and	Full-Scale IQ: (B) of -0.003, CI -0.006, 0.001, p= 0.064 Working Memory Index: (B) of -0.006, CI -0.009, -0.002, p<0.001. The	For each standard deviation increase in exposure (4.61pg/g) there is a 1.4% reduction in Full-Scale IQ and a 2.8% reduction in	Strengths• Direct assessment of chlorpyrifos levels using maternal serum and cord blood • Analysis using a continuous

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
	Checklist (CBCL).	blood (12% of subjects), levels were imputed from mothers' values. • Values for samples with non-detectable chlorpyrifos levels (N=115, 43%) were imputed by using assay-specific limit of detection (LOD) values to impute an approximate level.	passive smoking, alcohol and drug use during pregnancy, and residential pesticide use. Final models included prenatal environmental tobacco smoke (ETS) exposure, gender, ethnicity, gestational age at birth, quality of home care-taking environment, maternal education, and maternal IQ.	investigators articulated these results as showing that a 1 pg/g increase in chlorpyrifos exposure was associated with a 0.006 point decrease in the log-transformed Working Memory score and a 0.003 point decrease in the log-transformed Full-Scale IQ score. The investigators concluded that for each standard deviation increase in exposure (4.61pg/g) there is a 1.4% reduction in Full-Scale IQ and a 2.8% reduction in Working Memory.	Working Memory.	CPF level, which, in contrast to dichotomous CPF levels, provides a more meaningful look at potential threshold effects and dose-response trends.• The investigators rigorously evaluated their methods for imputing values for undetectable CPF levels which in the end, were validated.• The authors describe an elegant and methodologically sound statistical analysis, addressing many of the potential shortcomings of their exposure data and covariates.Weaknesses• The use of a single snapshot of prenatal chlorpyrifos exposure may not be an accurate surrogate for full prenatal exposure levels. • There is no control for exposure over the subsequent 7 years which may be critical, especially as the process of neurocognitive development is fluid and rapid during these early childhood years. • Possibility of that an increased awareness of the risks of pesticide exposures could disproportionately affect postnatal exposure behavior.• Complicating this analysis is the pervasive, non-specific nature of neurological effects and the difficulty in attributing causal pathways. • when closely reviewed, the 95% CI for Full Scale IQ for both techniques contain 0 (LASSO: -0.006, 0.001, p=0.064; fully-adjusted: -0.006, 0.001, p=0.048• The authors do

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
						not address the clinical relevance of the 1.4% and 2.8% reductions and how this may impact a child or his/her psychological or educational plans.
Article 9: Engel et al. (2011) Mt. Sinai (N=169)	Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III) at ages < 7 years; Wechsler-IV Intelligence Scale (verbal comprehension; perceptual reasoning, working memory, processing speed, full scale intelligence) at age 7-9 years	DEP: 24.7 nmol/L; Total DAP: 82 nmol/L (same as Engel 2007)	Maternal age, race/ethnicity, marital status, education, breast-feeding, child sex, alcohol, smoking, or drug use during pregnancy, maternal IQ, a score based on assessment of the home environment (HOME), season of urine collection, language spoken in the home, age at testing, examiner and urinary creatinine level.	At age 6-9 years, non-statistically significant reductions in full scale IQ, perceptual reasoning, verbal comprehension, working memory and processing speed with increasing DAP, more profound with DEP than DMP;	Used non-specific biomarker DEP/DAP; some evidence of effect modification by PON1 genotype	Limitations include use of non-specific markers of chlorpyrifos pesticide exposure (DAPs), use of only a single (third-trimester) urine sample, and the large proportion of loss to follow-up. The statistical analysis was largely appropriate. Imputing of missing data may affect precision of association estimates and result in attenuated effect estimates as a result of exposure measurement error, although these are offset by the further categorization of the exposure levels (at the median). Assessment and control for confounding were appropriate. However, confounding by unmeasured (and mis-measured) risk factors for abnormal growth that are related to pesticide exposure would bias the results in this study. Such factors may be related to socioeconomic status of the study participants, a cofactor which is difficult to define, no less measure in an epidemiologic study.
Article 14: Bouchard et al. (2011) CHAMACOS (N=329)	Wechsler-IV Intelligence Scale (verbal	Total DAPs (quintiles): Q1 (39 nmol/L); Q2 75 nmol/L; Q3 126	Maternal intelligence, measures of how stimulating the environment is, and known or	The authors observed evidence of an association between prenatal exposures to OP	Prenatal measures taken later half of pregnancy more significantly associated	This study has many strengths, the longitudinal design, the measurement of urinary DAP at

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
	comprehension; perceptual reasoning, working memory, processing speed, full scale intelligence) measured at age 7 years	nmol/L; Q4 221 nmol/L; Q5 508 nmol/L. Geometric mean DAP 131 nmol/L	suspected neurotoxins were measured prenatally. Maternal intelligence was assessed via the Peabody Picture Vocabulary Test (PPVT). To measure the quality and extent of stimulation available to a child in the home environment, the Infant-Toddler HOME (Home Observation for Measurement of the Environment) inventory was completed at the 6-month, 1, 2, 3.5, 5, and 7 year visits; known or suspected neurotoxicants, polybrominated diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs), p,p'-dichlorodiphenyltrichlorethane (DDT), p,p'-dichlorodiphenyltrichlorethylene (DDE), and lead.	pesticides as measured by urinary DAP metabolites in women during pregnancy, and decreased cognitive functioning in children at age 7.	intelligence than early; little evidence post-natal OP exposure associated with intelligence; 7 point reduction in full scale intelligence DAP Q5/Q1 (SS)	multiple time points and following children to age seven when tests of cognitive function are reportedly more reliable. The authors were able to adjust for or consider many factors related to cognitive function, such as prenatal exposure to other environmental agents, socioeconomic indicators, maternal intelligence and education, and child stimulation. The cohort had a relatively homogenous socioeconomic profile, reducing the potential for uncontrolled confounding.
Article 15: Whyatt et al. (2007) Columbia U. (N=102)	None	Geometric mean, 6.9 ± 17.0 ng/m^3; range < 0.4–171 ng/m^3. Personal air monitor: median 2.8 ng/m^3, mean 6.2 ± 11.1 ng/m^3, range < 0.4–83.4 ng/m^3	N/A	There was little within-home variability and no significant difference in air concentrations within homes over time ($p \geq 0.2$); between-home variability accounted for 88% of the variance in the indoor air levels of propoxur, 92% in chlorpyrifos, 94% in diazinon, and 62% in piperonyl butoxide ($p < 0.001$). Indoor and maternal personal air insecticide levels were highly correlated ($r = 0.7$–0.9, $p < 0.001$).	Indoor and maternal personal air insecticide levels were highly correlated ($r = 0.7$–0.9, $p < 0.001$).	Strengths: study design and exposure assessment techniques, Limitations: only those cohort participants enrolled after 2011 were included in the analysis (most likely due to the lack of serial data from the earlier years.)
Article 16: Whyatt et al. (2009) Columbia U. (N=102)	None	The limit of detection (LOD) of chlorpyrifos in blood samples was 0.5–1 pg/g plasma. The LOD of TCPy in urine samples was 0.26 ng/mL urine. The	N/A	Meconium TCPy concentrations were positively correlated with chlorpyrifos in maternal and cord blood ($r = 0.25$–0.33, $p < 0.05$) and with TCPy in maternal urine ($r = 0.31$, $p < 0.01$).	TCPy in maternal urine samples was not reliable, but the maternal and cord blood chlorpyrifos as well as the TCPy levels in meconium were reliable measures of exposure	Comprehensive exposure assessment including actual blood chlorpyrifos levels, the repeated sampling, and the environmental sampling. Weaknesses: only included

Author, Year, Sample Size	Outcome Assessed	Exposure Measurement	Potential confounders considered	Primary Result	Conclusions/Uncertainties	Strengths and Weaknesses
		LOD for TCPy in meconium was 0.2 ng based on a sample weighing 0.5 g. Exposure marker levels below the LOD were given a value of half the level of detection, and were then log10 transformed.				participants recruited in the post-cancellation period, use of nonparametric, rank-based statistics is appropriate but the large number of observations below the level of detection receiving equal rank, may be problematic; no dietary assessment