

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 ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

Association between Exposure to *p,p'*-DDT and Its Metabolite *p,p'*-DDE with Obesity: Integrated Systematic Review and Meta-Analysis

German Cano-Sancho, Andrew G. Salmon, Michele A. La Merrill

Table of Contents

SYSTEMATIC REVIEW PROTOCOL

Figure S1. Workflow for rating the quality and integration of evidence from human and animal evidence, and judgement of supporting in vivo and in vitro evidence for hazard identification conclusions.

Figure S2. Flow diagram for the study search and selection

Table S1. Relationship of confidence features and the main study designs (OHAT 2015b).

Table S2. Criteria to rate the risk of bias. To downgrade the confidence will integrate the risk of bias from each study providing relevant information for the health outcomes of interest.

Table S3. Criteria to rate the imprecision.

Table S4. Criteria to assess the inconsistency

Table S5. Summary for the confidence rating procedure

Table S6. Translation of confidence rating into level of evidence

Figure S3. Hazard identification scheme (OHAT 2015b).

Table S7. Proposed framework for systematic integration of supporting in vitro and in vivo evidence

Figure S4. Flow chart to present the results from the systematic review process.

Table S8. Prospective studies included in the meta-analysis to assess the association between the exposure to DDTs and obesity. Abbreviations: Gen, gender; M, male; F, female; MF, male and female combined, M/F male and female stratified. Cohorts: CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas; FLEHS I, first Flemish Environment and Health Study; RMCC, The Rhea Mother–Child Cohort; INMA-Sabadell, Infancia y Medio-Ambiente Child and Environment birth cohort; EYHS, Danish part of the European Youth Heart Study. Outcomes: BMI-z, body mass index z-score. Confounders: ACT, physical activity; ALC, alcohol; BIR-OR, birth origin; BIR-W, birth weight; BRE-FEE, breastfeeding; EDU, education; GES-AGE, gestational age; LIP, lipids; MAT-AGE, maternal age; MAT-BMI, maternal BMI; ORI, origin; PAR, parity; POL, pollutants; SES, socioeconomic status; SMO, maternal smoking. Un, units of exposure; lw, units normalized by lipid weight; ww, units in wet weight.

Table S9. Exposure levels in prospective studies included in the meta-analysis to assess the association between the exposure to p,p'-DDE and obesity. Studies with asterisk providing potentially overlapping information from the same cohort were only included in stratified analysis. Summary of effect estimates: (▲) statistically significant increase, (▼) statistically significant decrease, (∅) non-statistical significance. Abbreviations. CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas; FLEHS I, first Flemish Environment and Health Study; RMCC, The Rhea Mother–Child Cohort; INMA-Sabadell, Infancia y Medio-Ambiente Child and Environment birth cohort; EYHS, Danish part of the European Youth Heart Study. BMI, body mass index; GM, geometric mean; IQR, interquartile range.

Figure S5. Sensitivity analysis performing meta-analysis random-effect estimates omitting one study at the time.

Figure S6. Publication bias assessed by Egger's test and funnel plots. The funnel plots did not show asymmetric trend and the Egger's test did not provide statistically significant evidence of small-study effects for the sort of studies included in the meta-analysis.

Table S10. Risk of bias summary results. Classification: (++) definitively high risk of bias, (+) probably high risk of bias, (-) probably low risk of bias, (--) definitively low risk of bias. (*) Asterisk indicates a key risk of bias domain. T1, tier 1 according the NTP/OHAT tiered approach risk of bias tool approach (OHAT 2015a). Full instructions at "Section 6, Instructions to assess the risk of bias of human epidemiological studies".

Table S11. Risk of bias of Warner et al. 2014 . Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) Study, according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index.; TIME-US-B, time at United States at birth.

Table S12. Risk of bias of Delvaux et al. 2014, first Flemish Environment and Health Study (FLEHS I) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index; MAT-AGE, maternal age; SMO, maternal smoking; MAT-EDU, maternal education; LIP, serum lipids.

Table S13. Risk of bias of Hoyer et al. 2014 (INUENDO) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index; PAT-BMI, paternal body mass index; SMO, maternal smoking; ALC, maternal alcohol; EDU, maternal education; PAR, parity; MAT-AGE, maternal age; BRE-FEE, breast-feeding; ACT, physical activity; DIET, diet.

Table S14. Risk of bias of Vafeiadi et al. 2015, The Rhea Mother–Child Cohort (RMCC) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-TG, maternal triacylglycerides; MAT-CHOL, maternal cholesterol; MAT-AGE, maternal age; MAT-BMI, maternal body mass index; PAR, parity; EDU, education; SMO, maternal smoking; BRE-FEE, breast-feeding; BIR-W, weight at birth; GES-AGE, gestational age.

Table S15. Risk of bias of Cupul-Uicab et al. 2010, Tapachula study , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: SMO, maternal smoking; MAT-EDU, maternal education; MAT-BMI, maternal body mass index; GEST-AGE, gestational age; RECRUIT, recruitment.

Table S16. Risk of bias of Agay-Shay et al. 2015, Infancia y Medio-Ambiente Child and Environment birth cohort (INMA-Sabadell) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a).. Abbreviations: GEST-AGE, gestational age; BIR-W, weight at birth; ORI, origin; MAT-AGE, maternal age; MAT-BMI, maternal body mass index; MAT-W- Δ , maternal weight increase; SES, socioeconomic status; BREE-FEE, breast-feeding; SMO, maternal smoking.

Table S17. Risk of bias of Tang-Peronard et al. 2015, Danish part of the European Youth Heart Study (EYHS) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: SMO, smoking; ACT, physical activity; MAT-BMI, maternal body mass index; SEX, gender.

Table S18. Prospective studies excluded in the meta-analysis assessing associations between the exposure to DDTs and obesity. Abbreviations: Gen, gender; M, male; F, female; MF, male and female combined, M/F male and female stratified. Cohorts: CPP, Collaborative Perinatal Project; PIVUS, The Prospective Investigation of the Vasculature in Uppsala Seniors; AMICS-INMA-Menorca, Menorca Asthma Multicentre Infants Cohort Study– Infancia y Medio Ambiente. Outcomes: BMI, body mass index; WC, waist circumference; OVE, overweight; OBE, obesity. Statistic: \emptyset no statistically significant effect, \blacktriangle statistically significant increase.

Table S19. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and obesity. Obesity was defined as increased adiposity. Abbreviations: i.p. intraperitoneal injection; o.s. oral administration; PND, post-natal day.

Table S20. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and abnormal lipids. Abbreviations: BW, body weight; CHO, cholesterol; FFA, free fatty acids; GD, gestational day; g.t. gastrointestinal tube; HFD, high fat diet; i.p. intraperitoneal injection; i.t. intratraqueal; NCM, Northern Contaminant Mixture; NEFA, non-esterified fatty acids; NS, no specified; o.s. oral administration; POP, persistent organic pollutants; STD, standard diet; TAG, triacylglycerol.

Table S21. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and energy balance and adipokines. Abbreviations: BW, body weight; g.t., gastrointestinal tube ; i.p., intraperitoneal; i.v., intravenous; o.s., oral administration.

Table S22. Relationship of parts per million in diet to mg/kg body weight per day (JECFA 2000).

Table S23. Summary of study characteristics of in vitro evidence reporting associations of exposure to DDTs and adipogenesis, lipogenesis and markers of metabolic homeostasis. Abbreviations: ATGL, adipose triglyceride lipase; DMSO, dimethyl sulfoxide, PPAR, peroxisome proliferator-activated receptor; CEBP enhancer-binding protein; Lep, leptin; LpL, lipoprotein lipase; Insig1, Insulin-induced gene-1; Fabp4, Fatty acid binding protein 4; Fasn, fatty acid synthase; Srebfl, sterol regulatory element-binding protein 1c; Slc2a4, glucose transporter type 4.

Table S24. Summary table of risk of bias of in vivo studies. Classification: (++) definitively high risk of bias, (+) probably high risk of bias, (-) probably low risk of bias, (--) definitively low risk of bias. (*) Asterisk indicates a key risk of bias domain. T1 means tier 1 and T2, tier 2 according the NTP/OHAT (2015) tiered approach risk of bias tool approach. Risk of bias ratings according Navigation Guide instructions for non-human studies (Kousta et al. 2014), full instructions at “Section 7, Instructions to assess the risk of bias of in vivo studies”.

Table S25. Risk of bias of the study La Merrill et al. 2014. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S26. Risk of bias of the study Skinner et al. 2013. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S27. Risk of bias of the study Okazaki and Katayama 2003. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S28. Risk of bias of the study Okazaki and Katayama 2008. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S29. Risk of bias of the study Ishikawa et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S30. Risk of bias of the study Rodriguez-Alcala et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Table S31. Risk of bias of Howell et al. 2014. Risk of bias rating according Navigation Guide instructions for non-human studies (Koustas et al. 2014).

Table S32. Risk of bias of Howell et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Koustas et al. 2014).

Figure S7. Directed acyclic graph illustrating the causal associations between different exposures (direct and maternal), covariables and outcomes in the model where the levels of DDTs are adjusted by lipids (lipid weight basis). Covariables: ACT, physical activity; BIR-W, birth weight; BRE-FEE, breastfeeding; EDU, education; LIP, lipids; AGE, age; BMI, body mass index; PAR, parity; POPs: other persistent organic pollutants; RAC, race; SMO, smoking.

Figure S8. Summary of exposure levels from in vivo studies compared with human epidemiological studies to assess directness of levels. The internal doses of human studies are serum levels expressed in wet weight (black) and converted to wet weight from reported levels in lipid weight (red) using the conversion factor 1:129.8 wet weight:lipid weight (Lopez-Carrillo et al. 1999). Two approaches were assumed to compare the exposure levels of in vitro studies: level of cell culture exposure assuming accumulation of p,p'-DDT and p,p'-DDE in adipose tissues and using a ratio 1:129.8 serum:adipose tissue; and assuming the exposure in adipocytes without accumulation using a ratio 1:1 serum:adipose tissue.

Table S33. Summary of dose levels as reported by the supporting in vivo studies.

Figure S9. Summary of exposure levels from in vitro studies compared with human epidemiological studies to assess directness of levels. The internal doses of human studies are serum levels expressed in wet weight (black) and converted to wet weight from reported levels in lipid weight (red) using the conversion factor 1:129.8 wet weight:lipid weight (Lopez-Carrillo et al. 1999). Two approaches were assumed to compare the exposure levels of in vitro studies: level of cell culture exposure assuming accumulation of p,p'-DDT and p,p'-DDE in adipose tissues and using a ratio 1:129.8 serum:adipose tissue; and assuming the exposure in adipocytes without accumulation using a ratio 1:1 serum:adipose tissue.

HUMAN EVIDENCE

NON-HUMAN EVIDENCE

RATING AND INTEGRATION OF EVIDENCE

INSTRUCTIONS FOR RATING THE CONFIDENCE

DATA FORMS

INSTRUCTIONS TO ASSESS THE RISK OF BIAS OF HUMAN EPIDEMIOLOGICAL STUDIES

INSTRUCTIONS TO ASSESS THE RISK OF BIAS OF IN VIVO STUDIES

REFERENCES

1. SYSTEMATIC REVIEW PROTOCOL

1.1. INTRODUCTION

The Obesity Society defines obesity as a disease characterized by an excess of body fat, either total body fat or a particular depot of body fat, which increases the likelihood of comorbidities such as diabetes, hypertension, coronary heart disease, stroke, some cancers, obstructive sleep apnea or osteoarthritis (Allison et al. 2008; Arnold et al. 2015; Jokinen 2015). Obesity has been increasing in all countries, with prevalence doubling during the past three decades to reach global rates of 11 and 15% in 2014, for adult men and women, respectively, becoming a substantial public health concern worldwide (Ogden et al. 2014; WHO 2014). Excess caloric consumption and sedentary behavior are some of the risk factors traditionally identified as the main promoters of obesity and overweight, however the complex etiology of this condition involves multiple interrelated causes, such as genetic, social and environmental factors (Speakman and O'Rahilly 2012; WHO 2014).

The body of evidence for obesogenic effects of the pesticide dichlorodiphenyltrichloroethane (DDT) and its metabolite dichlorodiphenyldichloroethane (DDE) has increased notably in the last decade, with a particular focus on exposure during prenatal development. Technical DDT is a persistent organic pesticide mixture of three isoforms, *p,p'*-DDT, *o,p'*-DDT, and *p,p'*-DDD. In the present paper we use the term DDTs to identify the molecular family including these DDT isoforms and their metabolites (e.g. *p,p'*-DDE). The commercial formulation was widely used for the control of disease (e.g. malaria, typhus) vectors in agriculture from the mid-1940s to the late 20th century and is still manufactured in India for use primarily in India and Africa for control of malaria (Faroon and Harris 2002; Rogan and Chen 2005; UNEP 2010). Despite the Stockholm Convention implementation, use of DDT did not change substantially, representing the main insecticide, in terms of quantity used for vector control (71 % of total) in the above mentioned countries (van den Berg et al. 2012). Moreover, due to the extremely high persistence and lipophilicity of DDTs (considering the different isoforms and metabolites), internal exposure to this pesticide and its metabolites is ubiquitous in many countries, even decades after the ban was enforced (Rogan and Chen 2005; Smith 1999). The accepted adverse health effects of DDTs include impaired reproductive function, preterm birth, and it has recently been classified by the

International Agency for Research on Cancer (IARC) in the Group 2A, as probably carcinogenic to humans (ATSDR 2002; Beard 2006; Loomis et al. 2015).

A systematic review uses an explicit, pre-specified approach to identify, select, assess, and synthesize the data from studies in order to address a specific scientific or public health question. Systematic Review methods do not supplant the role of expert scientific judgment, public participation, or other existing processes used by OHAT and NTP in the evaluation of environmental substances. However, the systematic review methods are a major part of evidence-based decision making in terms of ensuring the collection of the most complete and reliable evidence to form the basis for decisions or conclusions (Johnson et al. 2014; Koustas et al. 2014; Lam et al. 2014; OHAT 2015b; Woodruff and Sutton 2014).

1.2. PROBLEM FORMULATION

1.2.1. Objectives of the search

The overall objective of this evaluation is to deliver hazard identification and characterization conclusions about whether DDT and derivatives are associated with obesity by integrating evidence from human, animal and considering support provided from mechanistic studies.

1.2.2. Search Question

Search question:

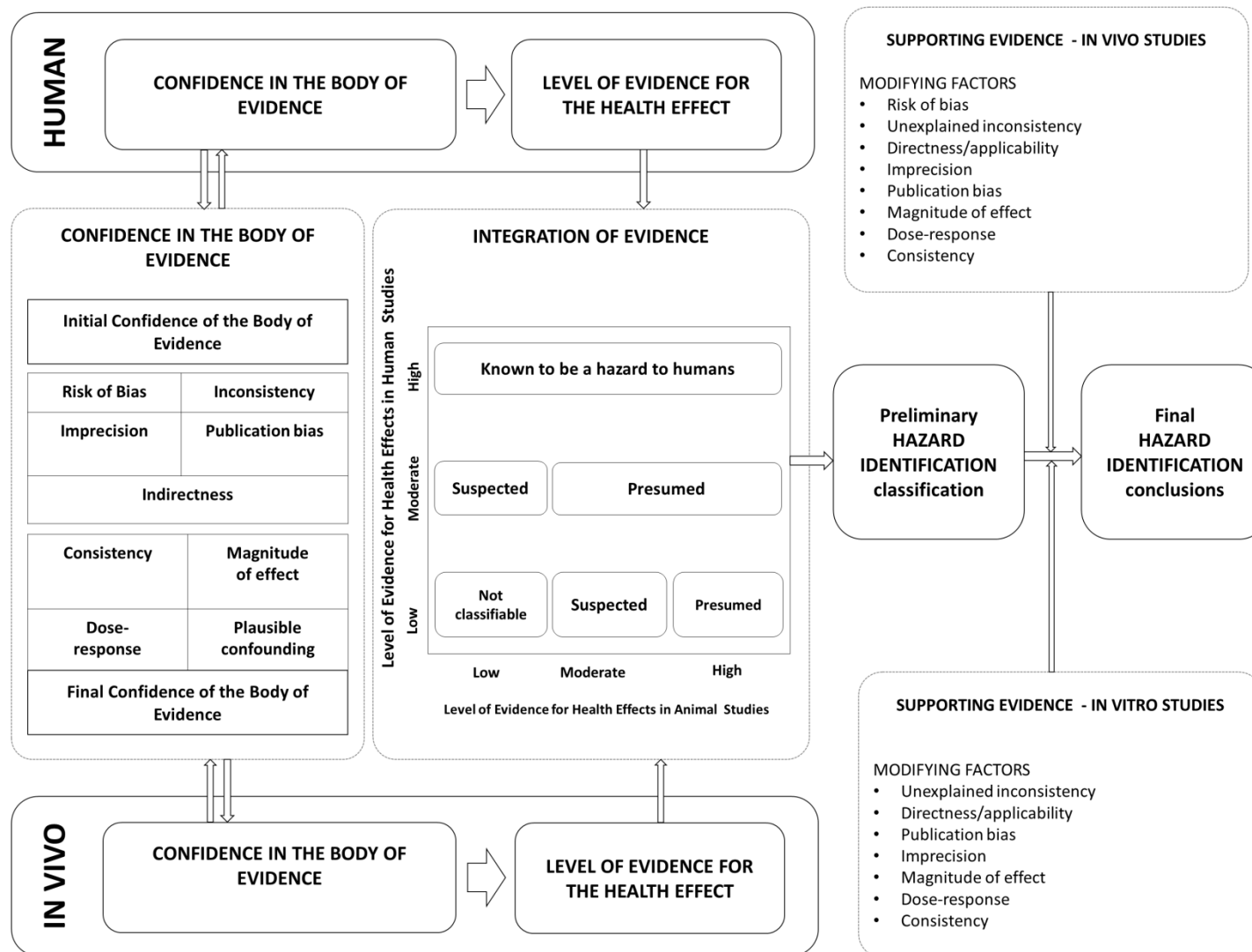
“Does exposure to DDT increase obesity in humans?”

1.3. SYSTEMATIC REVIEW METHODOLOGY

The SR methodology involves a systematic and well-documented process designed to gather the scientific evidence about obesogenic effects of DDT and its metabolites represented in the Figure S1. The search will be performed simultaneously in three different scientific databases, and additionally, studies published databases of grey literature will be also examined. Duplicate documents will be first eliminated from the pool of studies by means of reference manager software (ENDNOTE®). The library of studies will be uploaded to the specialized systematic review software DISTILLER SR where a second screening of duplicates will be performed. The entire process of selection and data extraction will be performed and

document in the DISTILLER SR platform. A first selection of studies will be performed considering only title and abstracts, examining the inclusion/exclusion criteria agreements. Exclusions will be recorded in a list of excluded documents, with a rationale description of exclusion reason. Doubtful studies, whose titles/abstracts do not provide enough information to decide, will be automatically moved to the next step to assess the full text. The full text of retained records will be gathered in pdf or paper version, and in case of unable full-text studies, the reason that limited the accessibility to the document will be clearly reported in the tracking document. The second level of selection based on the full-text will be also performed with DISTILLER SR[®] software. The data from the retained studies will be extracted using the specific data forms for each stream of evidence. The reviewer team will be formed by two reviewers and one external advisor. The selection, data extraction and data synthesis process will be performed by one reviewer (GCS) after checking the reproducibility, reliability and validity of outcomes by means of a full-duplicated pilot trial where two reviewers (GCS and MLM) will perform the entire process in a sub-sample of studies and will compare the outcomes. Results from the pilot trial must show no improvement of accuracy and reliability nor reduction of errors when the results from both reviewers are compared.

Figure S1. Workflow for rating the quality and integration of evidence from human and animal evidence, and judgement of supporting in vivo and in vitro evidence for hazard identification conclusions.



1.3.1. ELIGIBILITY CRITERIA

Populations

Epidemiological studies

Humans studied prospectively without restrictions on country, race, religion, sex will be considered. Cross-sectional studies will be excluded because some degree of reverse causality may be present due to the effect of adipose tissue on circulating DDT levels (Lee et al. 2012).

All ages and/or life-stage at exposure or outcome assessment will be included with exception of newborn (birth outcomes will be excluded).

In vivo studies

No restrictions on animal model, sex, age, life-stage at exposure or outcome assessment will be considered.

In vitro studies

No restrictions on cell lines and/or *in vitro* procedures will be considered.

Exposure

Epidemiological studies

Exposure to DDT and derivatives or isoforms based on administered dose or concentrations, environmental measures or indirect measures will be retained. The exposure must be measured individually using direct validated biomonitoring methods. We will exclude studies aiming to assess the therapeutic use of o,p'-DDD isoform, commercially known as mitotane or lysodren.

In vivo studies

Exposure to all type of DDT and derivatives or isoforms and their mixtures, including all range of concentrations, duration and routes of exposure will be retained. We will exclude studies including DDT in mixtures with other pollutants.

In vitro studies

Exposure to all type of DDT and derivatives or isoforms and mixtures, including all range of concentrations, duration and routes of exposure will be retained. We will exclude studies including DDT in mixtures with other pollutants.

Comparators

Epidemiological studies

Reference groups of population exposed at lower levels of DDTs than the rest of population groups will be considered.

In vivo studies

All vehicles in the control groups will be considered.

In vitro studies

All vehicles in the control groups will be considered.

Outcomes

Epidemiological studies

Primary outcome: body mass index (BMI), overweight and obesity.

Waist circumference or body fat distribution will be included as secondary outcomes.

In vivo studies

Primary outcome: adiposity. All measures of adiposity and fat weight will be considered.

Secondary outcomes: energy balance, abnormal lipids, other markers of metabolic homeostasis such as adipokines.

In vitro studies

Adipogenic differentiation, gene expression of metabolic regulators, adipokines.

Publication types

Only prospective epidemiological studies will be retained.

Reports must contain original data and being peer-reviewed.

All publication dates will be considered.

Articles not written in English will be excluded.

Conference papers will be excluded.

1.3.2. SEARCH

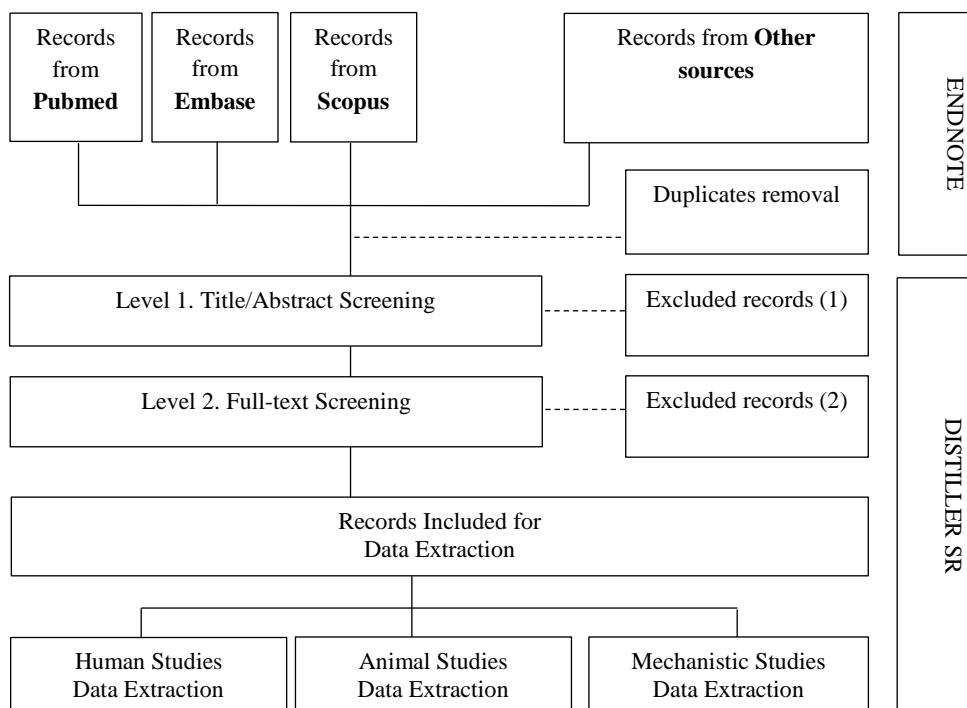
The search terms were extracted from published reviews and primary studies and identified by means of the PubMed Medical Subject Headings (MeSH). Additionally, we performed a supplementary search

through Embase Emtree database, open access databases and U.S. National Toxicology Program database to identify potential keywords.

No filters will be implemented during the search, including all publication years.

The search strategy will be built combining the main key elements (PECO elements) identified to answer the search question, nested through the Boolean operators AND/OR generating the search strings to implement in the scientific databases: Pubmed, Scopus and Embase (Figure S2).

Figure S2. Flow diagram for the study search and selection



Web Databases

- Pubmed / MeSH

<http://www.ncbi.nlm.nih.gov/pubmed>

<http://www.ncbi.nlm.nih.gov/mesh>

- Embase / Emtree.

<http://www.embase.com/#advancedSearch/default>

<http://www.embase.com/#emtreeSearch/default>

- Scopus

<http://www.scopus.com>

- Other sources
 - Secondary reference review of all extracted articles and the U.S. National Toxicology Program.
 - Open access databases.
 - Proceedings databases.
 - Governmental and grey literature databases
 - CalEPA Office of Environmental Health Hazard Assessment
 - ATSDR Toxicological profiles
 - European Chemicals Agency
 - European Food and Safety Authority
 - Health Canada
 - US National Toxicology
 - WHO Assessments

Search strings

Pubmed MeSH

("DDT"[MeSH] OR Dichlorodiphenyltrichloroethane [tiab] OR "Dichlorodiphenyl trichloroethane" [tiab] "50-29-3"[tiab] OR DDE [tiab] OR "Dichlorodiphenyl Dichloroethylene" [MeSH] OR DichlorodiphenylDichloroethylene [tiab] OR "72-55-9" OR "Dichlorodiphenyldichloroethane"[MeSH] OR "53-19-0") AND ("Diabetes Mellitus"[Mesh] OR diabetes [tiab] OR hyperglycemia [MeSH] OR "hypoglycemia"[MeSH] OR insulin [MeSH] OR insulin* [tiab] OR "blood glucose" [MeSH] OR "hemoglobin A, Glycosylated" [MeSH] OR gluconeogenesis [MeSH] OR "Glycolysis"[Mesh] OR glycolysis [tiab] OR "Glucose Transport Proteins, Facilitative"[Mesh] OR "metabolic syndrome x" [MeSH] OR "islets of Langerhans" [MeSH] OR "insulin-secreting cells"[MeSH] OR obesity [MeSH] OR obes* [tiab] OR overweight [MeSH] OR "body weight" [MeSH] "body mass index" [MeSH] OR "Waist-Hip Ratio"[MeSH] OR "Waist Circumference"[MeSH] OR "Skinfold Thickness"[MeSH] OR "Weight Gain"[MeSH] OR "Body Fat Distribution"[MeSH] OR "Adipose Tissue"[Mesh] OR Adipokines [MeSH] OR Adipogenesis [MeSH] adipokine*[tiab] OR adipocytokine*[tiab] OR adiponectin[mh] OR adiponectin*[tiab] OR ghrelin[mh] OR ghrelin[tiab] OR leptin[mh] OR leptin*[tiab] OR resistin[mh] OR resistin[tiab] OR Lipid metabolism[mh] OR lipogen*[tiab] OR lipid*[tiab] OR lipoprotein OR triacylglycerol OR triglyceride OR thermogenesis [MeSH] OR thermogen*[tiab])

Embase/EMTREE

('chlorphenotane'/exp OR '1,1,1 trichloro 2 (2 chlorophenyl) 2 (4 chlorophenyl)ethane'/exp OR '1,1 dichloro 2,2 bis(4 chlorophenyl)ethylene'/exp OR '1,1 dichloro 2,2 bis(4 chlorophenyl)ethane'/exp OR '50-29-3' OR '72-55-9' OR '53-19-0') AND ('diabetes mellitus'/exp OR 'hyperglycemia'/exp OR 'hypoglycemia'/exp OR 'insulin'/exp OR 'glucose blood level'/exp OR 'hemoglobin A1c'/exp OR 'gluconeogenesis'/exp OR 'glucose transporter'/exp OR 'metabolic syndrome X'/exp OR 'pancreas islet'/exp OR 'pancreas islet beta cell'/exp OR 'obesity'/exp OR obes*:ti:ab OR 'body mass'/exp OR 'body weight'/exp OR "body weight":ti:ab OR 'body weight disorder'/exp OR 'waist hip ratio'/exp OR 'waist circumference'/exp OR 'skinfold thickness'/exp OR 'weight gain'/exp OR 'body fat distribution'/exp OR 'adipose tissue'/exp OR 'adipogenesis'/exp OR 'adipocytokine'/exp OR adipocytokine*:ti:ab OR adipokine*:ti:ab OR 'adiponectin'/exp OR adiponectin*:ti:ab OR adipos*:ti:ab OR ghrelin/exp OR ghrelin:ti:ab OR leptin/exp OR leptin:ti:ab OR resistin/exp OR resistin:ti:ab OR lipogen*:ti:ab OR lipoprotein/exp OR lipoprotein*:ti:ab OR triacylglycerol/exp OR triacylglycerol:ti:ab OR triglyceride*:ti:ab OR 'thermogenesis and thermoregulation'/exp)

Scopus

TITLE-ABS-KEY (DDT OR Dichlorodiphenyltrichloroethane OR "Dichlorodiphenyl trichloroethane" OR "50-29-3" OR DDE OR DichlorodiphenylDichloroethylene OR "Dichlorodiphenyl Dichloroethylene" OR "72-55-9" OR DDD OR Dichlorodiphenyldichloroethane OR "Dichlorodiphenyl dichloroethane" OR "53-19-0")

AND TITLE-ABS-KEY (diabet* OR hyperglycemia OR hypoglycemia OR insulin OR "glucose blood level" OR "hemoglobin A1c" OR gluconeogenesis OR glycolysis OR "glucose transporter" OR "metabolic syndrome X" OR "pancreas islet" OR "pancreas islet beta cell" OR obes* OR "body mass" OR "body weight" OR "waist hip ratio" OR "waist circumference" OR "skinfold thickness" OR "weight gain" OR adipocyte* OR "body fat distribution" OR "adipose tissue" OR adipogen* OR adipocytokine* OR adipokine* OR adiponectin* OR adipos* OR ghrelin* OR leptin* OR resistin OR lipogen* OR lipoprotein* OR triacylglycerol* OR triglyceride* OR thermogen* OR thermoregulation)

1.3.3. SELECTION OF STUDIES

Selection of studies will be carried out in Distiller SR® software using sequential discriminatory process, based first on the Title/abstract agreement to the eligibility criteria, and second based on the full-text (Figure S2).

Step 1. Title/abstract screening

In this preliminary assessment, the title and abstract will be checked to match the inclusion and exclusion criteria. In case of conflict or unclear decision, the full-text review will be carried to clarify any decision. The excluded manuscripts will be confirmed and kept in a specific library.

Step 2. Full-text screening

The full document from each selected studies during the step 1 will be gathered in the available format. Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) are identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates. If necessary, study authors will be contacted to clarify any uncertainty about the independence of two or more articles. The studies included at this step will be classified in the different streams of evidence to proceed with the data extraction: human studies, *in vivo* studies and *in vitro* studies. Studies reporting simultaneously data from different streams of evidence will be included in each one. Only studies reporting results subjected to statistical analyses will be retained.

1.3.4. DATA EXTRACTION AND SYNTHESIS

The data will be extracted using data forms specifically designed for animal, human and *in vitro* studies, adapted from OHAT 2015 (See Annex 1). The data will be extracted by a main reviewer and checked by an additional external reviewer to guarantee the accuracy and reliability. Discrepancies and controversial issues will be discussed by the reviewer team, and external advice will be requested if it is required. Bullet questions about the quality will be included in the data form focusing to assess the risk of bias.

Rationale for selection of estimates (cohorts reported by multiple studies). When different publications are reporting outcomes from the same cohort, the publication reporting the greatest latency between exposure and outcome (oldest age at follow-up) will be retained. We will not collapse nor transform the effect estimates, being pooled as reported in the manuscript. In the pooled meta-analysis, we will use those estimates from combined gender when available, if not we will use the estimates from the different population sub-groups. Stratification analysis will be performed if more than two studies per category is available.

1.4. RATING THE BODY OF EVIDENCE

We will apply the NTP/OHAT framework, based on the GRADE guidelines, to rate the confidence with the body of evidence, translate to a level of evidence and integrate the different streams of evidence to deliver the hazard identification conclusions (OHAT 2015b). The overall work-flow process is illustrated in the Figure S1, considering two main bodies of evidence (human and *in vivo* studies) addressing the main health outcomes and we will consider a supplemental body of evidence with mechanistic data from *in vivo* and *in vitro* studies reporting mechanistic events and secondary outcomes related with obesity to support the preliminary classification. The quality and level of evidence will be evaluated independently for human and animal evidence, establishing an initial confidence rate and using a sequential process considering those factors that may affect (upgrading or downgrading) the confidence including the risk of bias, imprecision, publication bias, indirectness, magnitude, dose-response and plausible confounding.

The risk of bias will be evaluated by means of NTP-OHAT based risk of bias tools specifically designed for human epidemiological studies and animal studies and adapted for DDTs and obesity outcomes (Kousta et al. 2014; OHAT 2015b). We will not assess the risk of bias of *in vitro* studies due to the lack of risk of bias tools or guidance to assess the internal quality; however we considered the other factors to rate the confidence (Rooney et al. 2016). The rating process will be completed considering those upgrading and downgrading factors and balanced together to deliver a final rate.

The final confidence rating of each body of evidence (human and *in vitro*) will be translated to a level evidence and integrated using the hazard identification scheme to provide a preliminary classification of the chemical (“known”, “presumed”, “suspected” or “not classifiable” hazard for humans).

Two supplemental bodies of evidence will be established with supporting *in vivo* studies (reporting secondary outcomes) and *in vitro* studies and rated similarly to the main bodies of evidence to establish a final level of evidence. We will consider that a high level of supporting evidence may upgrade the preliminary classification, while a low level of evidence could downgrade. Moderate level of evidence will not modify the rate. We will judge together both bodies of supporting evidence integrating both levels of evidence to deliver a final decision to upgrade or downgrade the final rate.

Initial rating of confidence

The initial confidence rating will be determined by the main features determined by the study design:

1. The exposure to the substance is experimentally controlled
2. The exposure assessment demonstrates that exposures occurred prior to the development of the outcome (or concurrent with aggravation/amplification of an existing condition)
3. The outcome is assessed on the individual level (i.e., not through population aggregate data)
4. An appropriate comparison group is included in the study

Table S1. Relationship of confidence features and the main study designs (OHAT 2015b).

Study Design	Controlled Exposure	Exposure Prior to Outcome	Individual Outcome Data	Comparison Group Used	Initial Confidence Rating
Human controlled trial	likely	likely	likely	likely	high
Experimental animal	likely	likely	likely	likely	high
Cohort	unlikely	may or may not	likely	likely	low to moderate
Case-control	unlikely	may or may not	likely	likely	low to moderate
Cross-sectional	unlikely	unlikely	likely	likely	low
Ecologic	unlikely	may or may not	may or may not	likely	very low to moderate
Case series/report	unlikely	may or may not	likely	unlikely	very low to low

Factors downgrading the confidence

Risk of bias

The summary tables of risk of bias for each stream of evidence will be analyzed in order to analyze the overall consistency, direction, magnitude and sources of bias. Downgrading for risk of bias should reflect the entire body of studies; therefore, the decision to downgrade should be applied conservatively. The

decision to downgrade should be reserved for cases for which there is substantial risk of bias across most of the studies composing the body of evidence.

The NTP/OHAT’s risk of bias tiered approach considers some key elements or risk of bias domains of higher relevance to establish the classification criteria for each individual study. For observational human studies the key elements would typically include exposure assessment, outcome assessment, and confounding/selection.

Tier 1: A study must be rated as “definitely low” or “probably low” risk of bias for key elements AND have most other applicable items answered “definitely low” or “probably low” risk of bias.

Tier 2: Study meets neither the criteria for tiers.

Tier 3: A study must be rated as “definitely high” or “probably high” risk of bias for key elements AND have most other applicable items answered “definitely high” or “probably high” risk of bias.

Table S2. Criteria to rate the risk of bias. To downgrade the confidence will integrate the risk of bias from each study providing relevant information for the health outcomes of interest.

“Not likely”	Most information is from Tier 1 studies (low risk of bias for all key domains). Plausible bias unlikely to seriously alter the results
“Serious”	Most information is from Tier 1 and 2 studies. Plausible bias that raises some doubt about the results
“Very serious”	The proportion of information from Tier 3 studies at high risk of bias for all key domains is sufficient to affect the interpretation of results. Plausible bias that seriously weakens confidence in the results.

Imprecision

The assessment of the 95% confidence intervals is the primary method to assess imprecision by NTP/OHAT in agreement with the GRADE approach (Guyatt et al. 2011a).

Table S3. Criteria to rate the imprecision.

Not serious Rate:0	No or minimal indications of large standard deviations (i.e., $SD > \text{mean}$) For ratio measures (e.g., odds ratio, OR) the ratio of the upper to lower 95% CI for most studies (or meta-estimate) is < 10 ; or for absolute measures (e.g., percent control response) the absolute difference between the upper and lower 95% CI for most studies (or meta-estimate) is < 100
Serious Rate:-1	Does not clearly meet guidance for “not serious” or “very serious”
Very serious Rate:-2	Large standard deviations (i.e., $SD > \text{mean}$) For ratio measures (e.g., OR) the ratio of the upper to lower 95% CI for most studies (or meta-estimate) is ≥ 10 ; or for absolute measures (e.g., percent control response) the absolute difference between the upper and lower 95% CI for most studies (or meta-estimate) is ≥ 100

For continuous variables GRADE guidelines states that review authors should consider downrating for imprecision whenever there are sample sizes lower than 400. Similar to the procedure for dichotomous variables, it is possible to calculate the optimal information size (OIS) setting the α and β error (suggested at 0.05 and 0.2, respectively), mean difference (Δ) and selecting an appropriate standard deviation. On that basis, using the usual standards of α (0.05) and β (0.20), and an effect size of 0.2 standard deviations, representing a small effect, requires a total sample size of approximately 400 (200 per group) a sample size that may not be sufficient to ensure prognostic balance (Guyatt et al. 2011a).

Publication bias

The publication bias is defined by the “publication or non-publication of research findings, depending on the nature and direction of the results” (Higgins and Green 2011) and is assessed on the body of evidence, while the “selective outcome reporting” is assessed for each individual study during the risk of bias process (Guyatt et al. 2011d). Downgrading by publication bias is only considered when the concern to reduce the confidence is serious (OHAT 2015).

We considered the issues outlined by NTP/OHAT in agreement with GRADE to rate the publication bias:

- Early positive studies, particularly if small in size, are suspect.
- Publication bias should be suspected when studies are uniformly small, particularly when sponsored by industries, non-government organizations, or authors with conflicts of interest.
- Funnel plots, Egger’s regression, and trim and fill techniques can be used to visualize asymmetrical or symmetrical patterns of study results to help assess publication bias when adequate data for a specific outcome are available. Funnel plots and other approaches are less reliable when there are only a few studies.
- The identification of abstracts or other types of grey literature that do not appear as full-length articles within a reasonable time frame (around 3 to 4 years) can be another indication of publication bias.

Indirectness and applicability

To assess the extent of the directness and applicability, NTP/OHAT approach considers (1) relevance of the animal model to outcome of concern (2) directness of the endpoints to the primary health outcome(s) (3) nature of the exposure in human studies and route of administration in animal studies (4) duration of treatment in animal studies and length of time between exposure and outcome assessment in animal and prospective human studies. Similarly, GRADE group identifies four types of indirectness: differences in population (applicability), differences in interventions (applicability), differences in outcome measures (surrogate outcomes) and indirect comparisons (Guyatt et al. 2011b).

We outlined the following points to assess the directness in the present study:

- Differences in population (applicability) and relevance of the animal model to outcome of concern

Human studies. We may rate down for population differences if there is a compelling reason to justify the biology in the population of interest is so different of the population assessed and thus, the magnitude may differ substantially.

Animal studies. Studies conducted in mammalian model systems are assumed relevant for humans (i.e., not downgraded) unless compelling evidence to the contrary is identified during the course of the evaluation. The use of genetically modified mammalian models may be downgraded if the biology in such model may differ substantially to the human populations.

In vitro studies. The applicability of the cell model will be evaluated on the basis of the biological relevance in humans. For instance, the 3T3-L1 mice adipocyte is a cell model extensively implemented in the study of adipogenesis and obesity related outcomes in humans.

- Differences in outcome measures (surrogate outcomes) or directness of the endpoints to the primary health outcome(s).

The applicability of specific health outcomes or biological processes in non-human animal models is outlined in the PECO-based inclusion and exclusion criteria, with the most accepted relevant/interpretable outcomes considered “primary” and less direct measures, biomarkers of effect, or upstream measures of health outcome considered “secondary”.

- Nature of the exposure in human studies and route of administration in animal studies (OHAT 2015).

Human studies. Human studies are not downgraded for directness regardless of the exposure level or setting (e.g., general population, occupational settings, etc.). In NTP/OHAT’s process, the applicability of a given exposure scenario for reaching a “level of concern” for a certain subpopulation is considered after hazard identification.

Dose levels used in animal studies: There is no downgrading for dose level used in experimental animal studies because it is not considered as a factor under directness for the purposes of reaching confidence ratings for evidence of health effects. NTP/OHAT recognizes that the level of dose or exposure is an important factor when considering the relevance of study findings. In NTP/OHAT’s process, consideration of dose occurs after hazard identification as part of reaching a “level of concern” conclusion when the health effect is interpreted in the context of what is known regarding the extent and nature of human exposure.

Route of administration in animal studies: External dose comparisons used to reach level of concern conclusions need to consider internal dosimetry in animal models, which can vary based on route of administration, species, age, diet, and other cofactors. The most commonly used routes of administration (i.e., oral, dermal, inhalation, subcutaneous) are generally considered direct for the purposes of establishing confidence ratings.

Inconsistency

GRADE suggests rating down the quality of evidence if large inconsistency (heterogeneity) in study results remains after exploration of a priori hypotheses that might explain heterogeneity. Judgment of the

extent of heterogeneity is based on similarity of point estimates, extent of overlap of confidence intervals, and statistical criteria including tests of heterogeneity and I^2 . Apparent subgroup effects should be interpreted cautiously with attention to whether subgroup comparisons come from within rather than between studies; if tests of interaction generate low P-values; and whether subgroup effects are based on a small number of a priori hypotheses with a specified direction (Guyatt et al. 2011c). Inconsistency that can be explained, such as variability in study populations, would not be eligible for a downgrade. Potential sources of inconsistency across studies are explored, including consideration of population or animal model (e.g., cohort, species, strain, sex, life-stage at exposure and assessment); exposure or treatment duration, level, or timing relative to outcome; study methodology (e.g., route of administration, methodology used to measure health outcome); conflict of interest, and statistical power and risk of bias. Generally, there is no downgrade when identified sources of inconsistency can be attributed to study design features such as differences in species, timing of exposure, or health outcome assessment. There is no downgrade for inconsistency in cases where the evidence base consists of a single study. In this case, consistency is unknown and is documented as such in the summary of findings table (OHAT 2015b).

A useful statistic for quantifying inconsistency is

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100 \%$$

where Q is the chi-squared statistic and df is its degrees of. This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) (Higgins and Green 2011).

Thresholds for the interpretation of I^2 can be misleading, since the importance of inconsistency depends on several factors. A rough guide to interpretation is as follows:

- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity*;
- 50% to 90%: may represent substantial heterogeneity*;
- 75% to 100%: considerable heterogeneity*.

*The importance of the observed value of I^2 depends on (i) magnitude and direction of effects and (ii) strength of evidence for heterogeneity (e.g. P value from the chi-squared test, or a confidence interval for I^2) (Higgins and Green 2011).

Tau square (T^2 , τ^2 , τ^2): An estimate of the between-study variance in a random-effects meta-analysis. A τ^2 close to 0 would be strict homogeneity, and > 1 suggests the presence of substantial statistical heterogeneity (Higgins and Green 2011).

Table S4. Criteria to assess the inconsistency

“Not serious”	Point estimates similar Confidence intervals overlap Statistical heterogeneity is non-significant ($p \geq 0.1$) I ² of $\leq 50\%$
“Serious”	Point estimates vary Confidence intervals show minimal overlap Statistical heterogeneity has low p-value ($p \leq 0.1$) I ² of $> 50\%$ to 75%
“Very serious”	Point estimates vary widely Confidence intervals show minimal or no overlap Statistical heterogeneity has low p-value ($p \leq 0.1$) I ² of $> 75\%$

Factors upgrading the confidence

We considered three factors to upgrade the confidence with the main bodies of evidence as stated by GRADE: magnitude of effect, dose-response/gradient and plausible confounders (Guyatt et al. 2011e).

Magnitude

Large magnitude of effect will be considered to upgrade the confidence on the basis of NTP/OHAT and GRADE guidance. Large magnitude in human studies is based primarily on modeling studies that suggest confounding alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5) and very unlikely to explain associations with an RR greater than 5 (or less than 0.2) (Guyatt et al. 2011e; OHAT 2015b)

Dose-response

We considered upgrading the dose-response if there is enough evidence of monotonic and non-monotonic gradient (OHAT 2015).

Plausible confounding

Sources of potential plausible confounding, also known as “residual confounding” or “residual bias” in epidemiology need to be investigated specially among the human body of evidence based with observational studies.

The complex relationship between DDTs as well as the rest of lipophilic pollutants, serum lipids and obesity is not fully understood, and researchers commonly require of assumptions to formulate the causal models. The different approaches typically used for expressing serum levels of lipophilic compounds have pros and cons, which misused may trigger conflicting results (Li et al. 2013; O'Brien et al. 2015; Schisterman et al. 2005). The three most common approaches are 1) the serum exposure in lipid basis (ratio of pollutant serum levels by the triglycerides and cholesterol content), 2) including the serum lipid content as a covariate in the regression model and 3) using the unadjusted wet-weight values (Li et al.

2013). The first approach has been largely implemented in lipophilic pollutants arguing that such approach allows the comparison between populations or different tissue specimens (Schisterman et al. 2005). Nonetheless, the serum lipid content may be affected by the food composition, quantity and timing in case of non-fasting samples, as well as, physiological differences between genders (Phillips et al. 1989). Special attention is required when the chemical may be related causally with both, the health outcome and the serum lipid. For instance, the serum levels of DDTs has been positively correlated with triglycerides, LDL-C and HDL-C levels from the NHANES 99-06 cohort and associated with dyslipidemia in animal studies (Patel et al. 2012). Moreover a couple of simulation studies performed with PCBs concluded that the lipid standardization provided higher bias than using other approaches (Gaskins and Schisterman 2009; Schisterman et al. 2005). The adipose tissue is also well-known to act as storage of lipophilic pollutants, affected by several dynamic conditions such as fasting and/or weight loss, causing their mobilization to other compartments (La Merrill et al. 2013). Special attention will be put on the potential bias due to the different approaches implemented to manage the exposure levels and highlighted the necessity to carefully assess the possibility of residual bias due to the lipid standardization of exposure data.

Consistency

The consistency is outlined by the NTP/OHAT protocol as upgrading factor considering the consistency across animal studies, dissimilar populations and study types.

Types of consistency according NTP/OHAT approach (OHAT 2015b):

- *“across animal studies—consistent results reported in multiple experimental animal models or species”. There is no absolute definition of ‘consistency’ however finding the same direction of change in the same outcome in over two species would constitute sufficient evidence that a causal relationship has been established for IARC experimental evidence and consistency may be warranted (Preamble Part B Section 6).*
- *“across dissimilar populations—consistent results reported across populations (human or wildlife) that differ in factors such as time, location, and/or exposure “*
- *“across study types—consistent results reported from studies with different design features, e.g., between prospective cohort and case-control human studies or between chronic and multigenerational animal studies “*

Final rate of confidence

The final rate of confidence will be based on the judgement of all downgrading and upgrading factors over the initial rating. The final rates for each body of evidence are high confidence, moderate confidence and low confidence.

Table S5. Summary for the confidence rating procedure

Initial rating	Factors reducing confidence					Factors increasing confidence				Final rating
	Risk of bias	Unexplained inconsistency	Indirectness	Imprecision	Publication bias	Magnitude	Dose Response	Residual confounding	Consistency	
High			Serious /Not serious	Serious /Not serious						High
Moderate	Likely /Unlikely	Serious /Not serious			Likely /Unlikely	Large /Not large	Large /Not large	Large /Not large	Large /Not large	Moderate
Low										Low

Translation of confidence in the body of evidence into level of evidence for the health effect

We used the descriptors proposed by NTP/OHAT to translate the level of confidence into the level of evidence for the health effect for each stream of evidence considering the confidence in the body of evidence and direction of the health effect.

Five descriptors are used by NTP/OHAT to defined the levels of evidence:

- a. High level of evidence. There is moderate confidence in the body of evidence for an association between exposure to the substance and the health outcome(s)*
- b. Moderate level of evidence. There is low confidence in the body of evidence for an association between exposure to the substance and the health outcome(s), or no data are available.*
- c. Low level of evidence. There is low confidence in the body of evidence for an association between exposure to the substance and the health outcome(s), or no data are available.*
- d. Evidence of no health effect. There is high confidence in the body of evidence that exposure to the substance is not associated with the health outcome(s).*
- e. Inadequate evidence. There is insufficient evidence available to assess if the exposure to the substance is associated with the health outcome(s).*

The direction or nature of the effect was considered in the translation process as following:

Table S6. Translation of confidence rating into level of evidence

Confidence in the body of evidence	Direction of the effect (Health effect)	Level of evidence for the health effect
High	⇒	High
Moderate	⇒	Moderate
Low	⇒	Low
Very low or no evidence	⇒	Inadequate
Confidence in the body of evidence	Direction of the effect (No Health effect)	Level of evidence for the health effect
High	⇒	Evidence of no health effect
Moderate	⇒	Inadequate
Low	⇒	Inadequate
Very low or no evidence	⇒	Inadequate

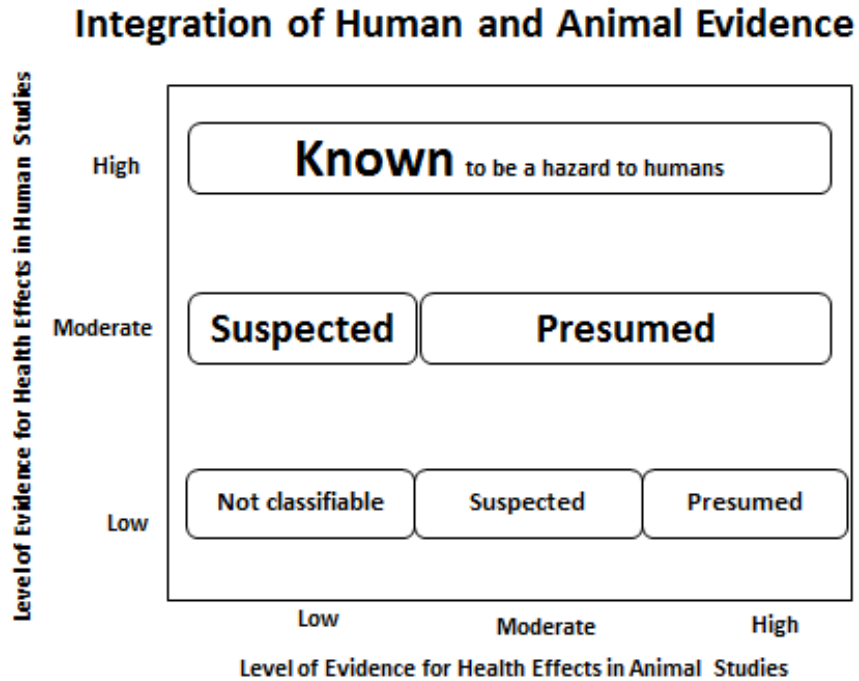
1.5. INTEGRATION OF EVIDENCE AND HAZARD IDENTIFICATION CONCLUSIONS

We integrated the human and animal main bodies of evidence considering the level of evidence established for each stream of evidence to deliver a preliminary hazard identification classification using the conceptual framework proposed by NTP/OHAT (OHAT 2015b) which is also in agreement of the IARC scheme for evidence integration, see Figure S3 (IARC 2006).

The hazard identification classes considered were:

- a. Known to be a hazard to humans
- b. Presumed to be a hazard to humans
- c. Suspected to be a hazard to humans
- d. Not classifiable as a hazard to humans
- e. Not identified as a hazard to humans

Figure S3. Hazard identification scheme (OHAT 2015b).



We considered the supporting bodies of evidence from *in vivo* and *in vitro* studies to judge upgrading or downgrading the preliminary hazard identification conclusions. The procedures for integration of mechanistic data in hazard identification evidence have not been well defined yet. However the growing literature supporting the biological plausibility of toxic effects requires of systematic and standardized approaches to be included in the hazard identification settings.

We considered that a high level of supporting evidence may upgrade the preliminary classification, while a low level of evidence could downgrade. Moderate level of evidence does not modify the rate. We judged together both bodies of supporting evidence integrating both levels of evidence to deliver a final decision to upgrade or downgrade the final rate.

Table S7. Proposed framework for systematic integration of supporting in vitro and in vivo evidence

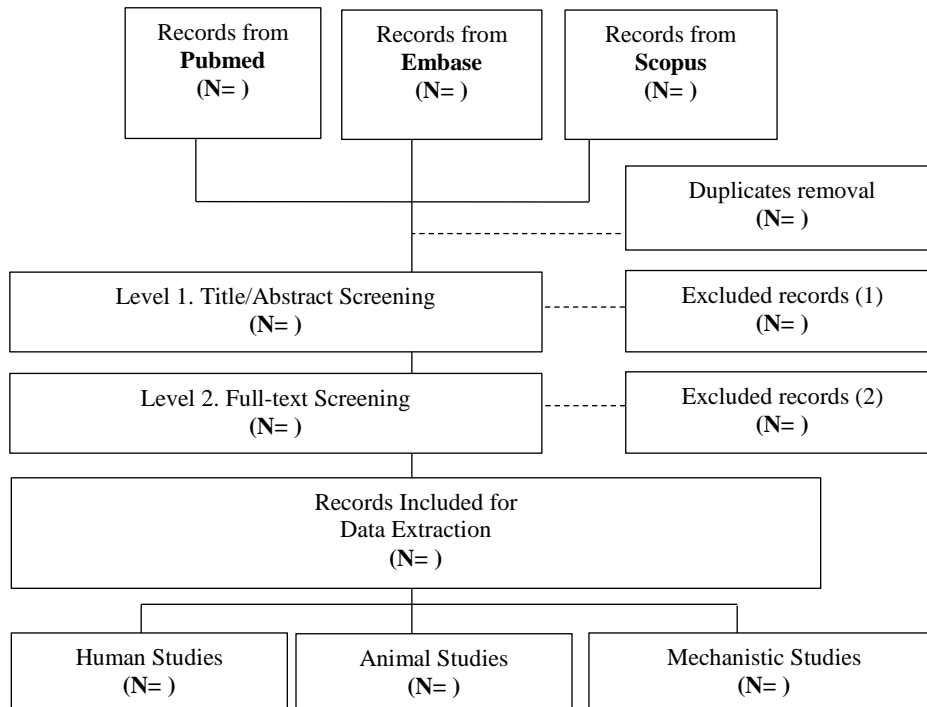
Preliminary classification	Level of supporting evidence	Direction	Final classification
Known	High	NA	Known
	Moderate	No upgrading/downgrading	Known
	Low	Downgrading	Presumed
Presumed	High	Upgrading	Known
	Moderate	No upgrading/downgrading	Presumed
	Low	Downgrading	Suspected
Suspected	High	Upgrading	Presumed
	Moderate	No upgrading/downgrading	Suspected
	Low	Downgrading	Not classifiable

1.6. PRESENTATION OF RESULTS AND SYNTHESIS OF DATA

Results from the systematic review will be presented graphically in a flow chart as presented in the Figure S4. The specific comments on the selection process decision will be attached in an annex justifying every decision.

Synthesis of data will be conducted quantitatively applying meta-analysis techniques whenever the quality of data allows. The meta-analysis will not be performed in those cases when the combination is not meaningful or the risk of bias from the studies is too high. Random-effects meta-analyses or fixed-effect meta-analyses will be implemented depending on the results, and the criteria will be supported by the judgement of a statistician.

Figure S4. Flow chart to present the results from the systematic review process.



2. HUMAN EVIDENCE

Table S8. Prospective studies included in the meta-analysis to assess the association between the exposure to DDTs and obesity.

Abbreviations: Gen, gender; M, male; F, female; MF, male and female combined, M/F male and female stratified. Cohorts: CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas; FLEHS I, first Flemish Environment and Health Study; RMCC, The Rhea Mother–Child Cohort; INMA-Sabadell, Infancia y Medio-Ambiente Child and Environment birth cohort; EYHS, Danish part of the European Youth Heart Study. Outcomes: BMI-z, body mass index z-score. Confounders: ACT, physical activity; ALC, alcohol; BIR-OR, birth origin; BIR-W, birth weight; BRE-FEE, breastfeeding; EDU, education; GES-AGE, gestational age; LIP, lipids; MAT-AGE, maternal age; MAT-BMI, maternal BMI; ORI, origin; PAR, parity; POL, pollutants; SES, socioeconomic status; SMO, maternal smoking. Un, units of exposure; lw, units normalized by lipid weight; ww, units in wet weight.

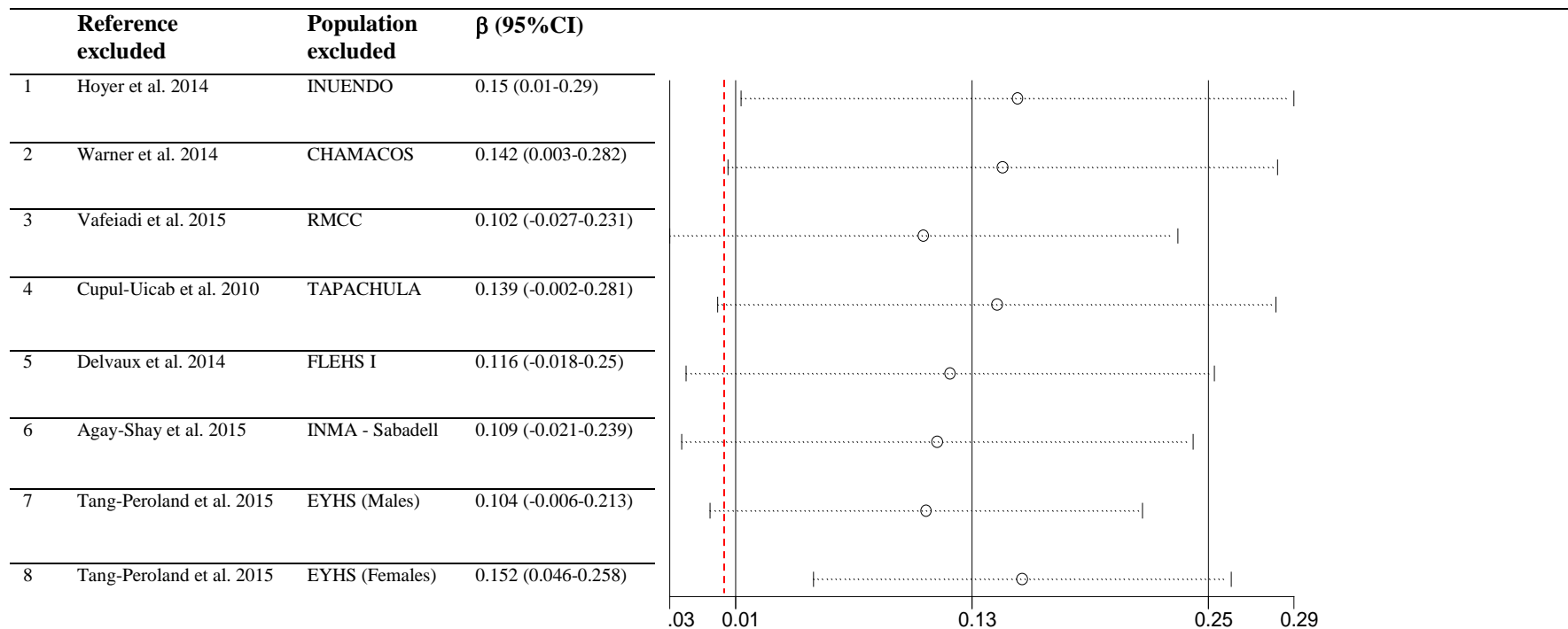
Cohort	Ref	Dates	Country	Gen	Age ^a	N	Outcome	Confounders	Exposure assessment	Un	Statistic
CHAMACOS	(Warner et al. 2014)	99-00	US	M/F/MF	9	261	BMI-z	MAT-BMI, TIME-US-B	Maternal Serum p,p'-DDE	lw	Beta
FLEHS I	(Delvaux et al. 2014)	02-11	Belgium	MF	7-8	114	BMI-z	MAT-BMI, MAT-AGE, SMO, MAT-EDU, LIP, AGE, SEX	Cord blood p,p'-DDE	ww	Beta
INUENDO	(Hoyer et al. 2014)	02-10	Greenland Poland Ukraine	MF	5-9	525 92 492	BMI-z	MAT-BMI, PAT-BMI, SMO, ALC, EDU, PAR, MAT-AGE, BRE-FE, ACT, DIET	Maternal Serum p,p'-DDE	lw	Beta
RMCC	(Vafeiadis et al. 2015)	07-11	Greece	MF	4	689	BMI-z	MAT-TG, MAT-CHOL, MAT-AGE, MAT-BMI, PAR, EDU, SMO, BRE-FEE, SEX, BIR-W, GES-AGE, POL	Maternal Serum p,p'-DDE	ww	Beta
INMA-Sabadell	(Agay-Shay et al. 2015)	04-06	Spain	MF	7	470	BMI-z	SEX, GEST-AGE, BIR-W, ORI, MAT-AGE, MAT-BMI, MAT-W-Δ, SES, BREE-FEE, SMO, ACT, DIET	Maternal Serum p,p'-DDE	lw	Beta
TAPACHULA	(Cupul-Uicab et al. 2010)	02-03	Mexico	M	0.4-2.5	788	BMI-z	SMO, MAT-EDU, MAT-BMI, GEST-AGE, RECRUIT	Maternal Serum p,p'-DDE	lw	Beta
EYHS	(Tang-Peronard et al. 2015)	1997-2009	Denmark	M/F	20-22	106/130	BMI-z	SMO, SES, ACT, MAT-BMI, SEX	Child serum levels p,p'-DDE	lw	Beta

^a Age at follow-up clinical evaluation (years)

Table S9. Exposure levels in prospective studies included in the meta-analysis to assess the association between the exposure to *p,p'*-DDE and obesity. Studies with asterisk providing potentially overlapping information from the same cohort were only included in stratified analysis. Summary of effect estimates: (▲) statistically significant increase, (▼) statistically significant decrease, (∅) non-statistical significance. Abbreviations. CHAMACOS, Center for the Health Assessment of Mothers and Children of Salinas; FLEHS I, first Flemish Environment and Health Study; RMCC, The Rhea Mother–Child Cohort; INMA-Sabadell, Infancia y Medio-Ambiente Child and Environment birth cohort; EYHS, Danish part of the European Youth Heart Study. BMI, body mass index; GM, geometric mean; IQR, interquartile range.

Cohort	Reference	Dates	Exposure levels as reported	Exposure assessment	Summary Effect Estimates
CHAMACOS	(Warner et al. 2014)	99-00	<u>Median (IQR) (ng/g lipid)</u> p,p'-DDE; 1,104 (613–2,710)	Maternal Serum p,p'-DDE	BMI-z p,p'-DDE ♂∅-∅♀
FLEHS I	(Delvaux et al. 2014)	02-11	<u>P50 (P25, P75) p,p'-DDE (µg/L)</u> 0.24 (0.13, 0.44)	Cord blood p,p'-DDE	▲ BMI-z, WC IQR ♀
INUENDO-Greenland	(Hoyer et al. 2014)	02-10	<u>p,p'-DDE (ng/g lipid)</u> Tertile low; 5.3–205.2 Tertile medium; 205.3–439.3 Tertile high; 439.4–3122.0	Serum p,p'-DDE	∅ BMI-z
INUENDO-Poland	(Hoyer et al. 2014)	02-10	<u>p,p'-DDE (ng/g lipid)</u> Tertile low; 88.1–302.8 Tertile medium; 302.9–471.2 Tertile high; 471.3–1750.1	Serum p,p'-DDE	∅ BMI-z
INUENDO-Ukraine	(Hoyer et al. 2014)	02-10	<u>p,p'-DDE (ng/g lipid)</u> Tertile low; 147.1–487.8 Tertile medium; 487.9–790.4 Tertile high; 790.5–4835.6	Serum p,p'-DDE	∅ BMI-z
RMCC	(Vafeiadi et al. 2015)	07-11	<u>GM (95% CI) (ng/L)</u> 2036.2 (1910.2, 2170.5) Percentile 25; 1187.6 Percentile 50; 1981.2 Percentile 75; 3514.1	Serum p,p'-DDE	▲ BMI-z ▲ Obesity
INMA-Sabadell	(Agay-Shay et al. 2015)	04-06	<u>Mean (95% CI) GM (ng/g lipid)</u> 236.4 (152.3, 320.5) 126.3	Maternal serum p,p'-DDE	▲ T2, T3 BMI-z
Tapachula, Mexico	(Cupul-Uicab et al. 2010)	02-03	<u>p,p'-DDE (µg/g)</u> < 3.01 3.01-6.00 6.01-9.00 >9.00	Serum p,p'-DDE	∅ BMI-z
EYHS	(Tang-Peronard et al. 2015)	97-09	<u>Median (Range) (µg/g lipid)</u> ♀ 0.04 (0.01–0.72) ♂ 0.20 (0.02–1.15)	Serum p,p'-DDE	BMI-z T2∅♂-∅♀ T3▲♂-▼♀

Figure S5. Sensitivity analysis performing meta-analysis random-effect estimates omitting one study at the time.



Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Figure S6. Publication bias assessed by Egger's test and funnel plots. The funnel plots did not show asymmetric trend and the Egger's test did not provide statistically significant evidence of small-study effects for the sort of studies included in the meta-analysis.

Begg's test

adj. Kendall's Score (P-Q) = 9
 Std. Dev. Of Score = 8.02(corrected for ties)
 N=8
 z=1.12
 P>|z|0.262
 z=1 (continuity corrected)
 P>|z| 0.319 (continuity corrected)

Egger's test

Std_Eff	Coef.	Std.	Err.	t	P> t	[95%CI]
slope	-0.142	0.237	-0.600	0.571	-0.722	0.438
bias	2.051	1.821	1.130	0.303	-2.405	6.507

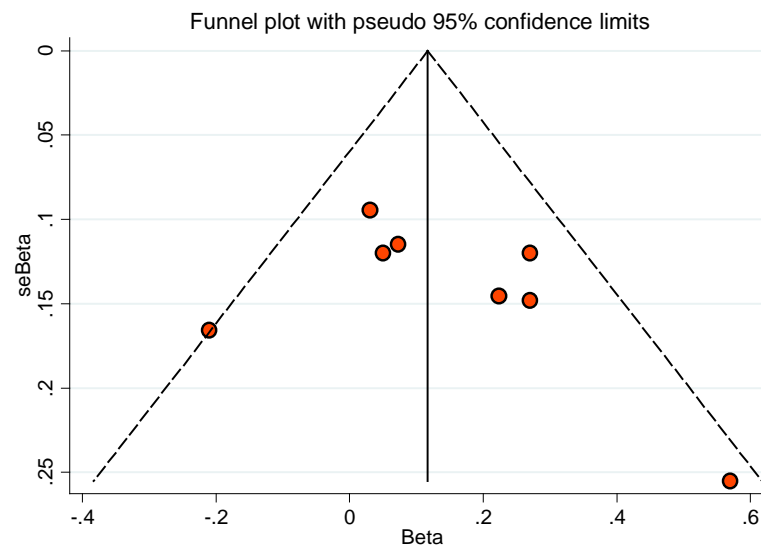


Table S10. Risk of bias summary results. Classification: (++) definitively high risk of bias, (+) probably high risk of bias, (-) probably low risk of bias, (--) definitively low risk of bias. (*) Asterisk indicates a key risk of bias domain. T1, tier 1 according the NTP/OHAT tiered approach risk of bias tool approach (OHAT 2015a). Full instructions at “Section 6, Instructions to assess the risk of bias of human epidemiological studies”.

Bias	Warner et al. 2014	Delvaux et al. 2014	Hoyer et al. 2014	Vafeiadi et al. 2015	Cupul-Uicab et al. 2010	Agay-Shay et al. 2015	Tang-Péronard et al. 2015
CONFOUNDING BIAS. * Did the study design or analysis account for important confounding and modifying variables?	(-)	(-)	(-)	(-)	(-)	(-)	(-)
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	(+)	(-)	(-)	(-)	(-)	(--)	(-)
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	(-)	(--)	(--)	(-)	(-)	(-)	(+)
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	(-)	(--)	(-)	(-)	(--)	(-)	(-)
Can we be confident in the exposure characterization? *	(--)	(--)	(-)	(--)	(--)	(-)	(--)
Can we be confident in the outcome assessment? *	(--)	(--)	(-)	(--)	(--)	(--)	(--)
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	(--)	(--)	(--)	(--)	(--)	(--)	(--)
CONFLICT OF INTEREST	(--)	(-)	(--)	(--)	(--)	(--)	(--)
Summary Tiered Classification	<i>T1</i>	<i>T1</i>	<i>T1</i>	<i>T1</i>	<i>T1</i>	<i>T1</i>	<i>T1</i>

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S11. Risk of bias of Warner et al. 2014 . Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) Study, according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index.; TIME-US-B, time at United States at birth.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS. Did the study design or analysis account for important confounding and modifying variables?	Probably low	Confounders included in the model are: MAT-BMI, TIME-US-B and stratified by gender
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably high	Other potential obesogenic chemicals not assessed and/or controlled, and setting of occupational exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Probably low	<i>“The children included in the analysis did not differ significantly from those who were excluded because of missing maternal serum DDT and DDE levels or 9-year anthropometric data (data not shown)”</i>
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Probably low	The authors did not report blinding but the study design prevents knowledge of exposure groups.
Can we be confident in the exposure characterization?	Definitively low	Exposure levels in wet weight and lipid weight. Used a validated isotope dilution gas chromatography-high resolution mass spectrometry methods reported at Barr et al. we be confident in the exposure characterization 2003. Mean levels of detection for o,p'-DDT, p,p'-DDT, and p,p'-DDE were 1.3 (standard deviation (SD), 0.7), 1.5 (SD, 0.8), and 2.9 (SD,1.5) pg/g serum, respectively.
Can we be confident in the outcome assessment?	Definitively low	<i>“Children were weighed and measured by trained research staff at each visit. At the 9-year visit, we measured barefoot standing height to the nearest 0.1 cm using a stadiometer and standing weight to the nearest 0.1 kg using a bioimpedance scale (Tanita TBF-300A Body Composition Analyzer, Tanita Corporation of America, Inc., Arlington Heights, Illinois) that also measured percentage of body fat using “foot-to-foot” bioimpedance technology. Height and waist circumference were measured in triplicate and averaged for analysis”</i>
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study’s specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	None declared. This work was supported by grants from the National Institute of Environmental Health Sciences at the National Institutes of Health, the National Institute for Occupational Safety and Health and the US Environmental Protection Agency

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S12. Risk of bias of Delvaux et al. 2014, first Flemish Environment and Health Study (FLEHS I) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index; MAT-AGE, maternal age; SMO, maternal smoking; MAT-EDU, maternal education; LIP, serum lipids.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	Confounding factors included in the model: MAT-BMI, MAT-AGE, SMO, MAT-EDU, LIP, AGE, SEX (stratified by gender)
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably low	Other potential obesogenic compounds measured but not controlled in multipollutant models. Not from a population of high occupational or acutely contaminating exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Definitively low	Missing data adequately addressed
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Definitively low	<i>The study nurses had no access to the prenatal exposure data when performing the measurements.</i>
Can we be confident in the exposure characterization?	Definitely low	Cord blood p,p'-DDE, units in wet weight. Robust method with additional supporting information. Gas chromatography-electron capture detection using the method of Gomara et al.(2002). The LOD for all chlorinated compounds was 0.02 µg/L.
Can we be confident in the outcome assessment?	Definitively low	Anthropometric data measured by the study nurses
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study's specified outcomes were adequately reported.
CONFLICT OF INTEREST	Probably low	The studies of the Flemish Center of Expertise on Environment and Health were commissioned, financed and steered by the Ministry of the Flemish Community

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S13. Risk of bias of Hoyer et al. 2014 (INUENDO) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-BMI, maternal body mass index; PAT-BMI, paternal body mass index; SMO, maternal smoking; ALC, maternal alcohol; EDU, maternal education; PAR, parity; MAT-AGE, maternal age; BRE-FEE, breast-feeding; ACT, physical activity; DIET, diet.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	Confounding factors included in the model: MAT-BMI, PAT-BMI, SMO, ALC, EDU, PAR, MAT-AGE, BRE-FEE, ACT, DIET, SEX
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably low	Measured levels of PCB-153 but not controlled in the model. Not from a population of high occupational or acutely contaminating exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Definitively low	Missing data adequately addressed. <i>“Missing information. The number of missing values on height, weight and covariates ranged from 0 to 27%. As complete case analysis may lead to selection bias, we addressed the missing information problem, using chained multiple imputation allowing us to include participants with incomplete data in the statistical analyses.”</i>
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Probably low	The authors did not report blinding but the study design prevents knowledge of exposure groups.
Can we be confident in the exposure characterization?	Definitively low	Serum p,p'-DDE in lipid weight. Robust method referring extra information to external citation (Jonsson et al. 2005). The sera were analyzed by gas chromatography-mass-spectrometry following solid phase extraction. Some measures of quality reported in the main text. To estimate the postnatal cumulative contribution of the compounds for the first 12 months after birth, a toxicokinetic model developed by Verner et al. 2013
Can we be confident in the outcome assessment?	Probably low	Some participants provided measurements by telephone-interview. <i>“All measurements were performed by the interviewer except for those who were telephone-interviewed.”</i> Telephone interviews were performed when families lived in remote areas of Greenland (n = 130) or had moved to Denmark (n = 34). Also, in Greenland, a proportion of the questionnaires was filled in by the parents without an interview.”
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study’s specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	None declared. The CLEAR (Climate change, environmental contaminants and reproductive health, http://www.inuendo.dk/clear) and INUENDO (Biopersistent organochlorines in diet and human fertility) studies were funded by the European Commission’s 7th and 5 th Framework Programmes, respectively

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S14. Risk of bias of Vafeiadi et al. 2015, The Rhea Mother–Child Cohort (RMCC) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: MAT-TG, maternal triacylglycerides; MAT-CHOL, maternal cholesterol; MAT-AGE, maternal age; MAT-BMI, maternal body mass index; PAR, parity; EDU, education; SMO, maternal smoking; BRE-FEE, breast-feeding; BIR-W, weight at birth; GES-AGE, gestational age.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	Confounding factors included in the model: MAT-TG, MAT-CHOL, MAT-AGE, MAT-BMI, PAR, EDU, SMO, BRE-FEE, SEX, BIR-W, GES-AGE.
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably low	They estimated associations using separate models for DDE, HCB, and the sum of PCBs. Not from a population of high occupational or acutely contaminating exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Probably low	The authors did not mention missing or incomplete data
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Probably low	The authors did not report blinding but the study design prevents knowledge of exposure groups.
Can we be confident in the exposure characterization?	Definitively low	Robust method, serum levels in wet weight. The POP analyses were performed in the National Institute for Health and Welfare, Chemical Exposure Unit, Kuopio, Finland with an Agilent 7000B gas chromatograph triple quadrupole mass spectrometer (GC-MS/MS). Pretreatment of serum samples for GCMS/MS analysis has been described elsewhere (Koponen et al. 2013).
Can we be confident in the outcome assessment?	Definitively low	Used standardized and reliable methods. No specified who performed the anthropometric measurements.
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study's specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	None declared. The Rhea project was financially supported by European projects and the Greek Ministry of Health

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S15. Risk of bias of Cupul-Uicab et al. 2010, Tapachula study , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: SMO, maternal smoking; MAT-EDU, maternal education; MAT-BMI, maternal body mass index; GEST-AGE, gestational age; RECRUIT, recruitment.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	Confounding factors included in the model: SMO, MAT-EDU, MAT-BMI, GEST-AGE, RECRUIT.
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably low	No information about levels of other potentially obesogenic pollutants. Not from a population of high occupational or acutely contaminating exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Probably low	No relevant missing data among followed subjects
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Definitively low	The present study was double blind, since neither interviewers nor participants knew the DDT or DDE levels.
Can we be confident in the exposure characterization?	Definitively low	Serum p,p'-DDE, units in lipid weight. Robust methodology, reporting quality control measures. DDE and DDT were quantified after solid phase extraction, using gas chromatography with mass spectrometry detection (Saady and Poklis, 1990; Smith, 1991).
Can we be confident in the outcome assessment?	Definitively low	No details on who performed the anthropometric measurements but used standardized procedures
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	The pre-specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	None declared. Governmental source of funding.

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S16. Risk of bias of Agay-Shay et al. 2015, Infancia y Medio-Ambiente Child and Environment birth cohort (INMA-Sabadell) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a).. Abbreviations: GEST-AGE, gestational age; BIR-W, weight at birth; ORI, origin; MAT-AGE, maternal age; MAT-BMI, maternal body mass index; MAT-W-Δ, maternal weight increase; SES, socioeconomic status; BREE-FEE, breast-feeding; SMO, maternal smoking.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	The authors adjusted for the main confounders like gender, maternal BMI or maternal smoking. Confounding factors included in the model: SEX, GEST-AGE, BIR-W, ORI, MAT-AGE, MAT-BMI, MAT-W-Δ, SES, BREE-FEE, SMO
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Definitively low	The authors applied a single-pollutant and a multi-pollutant model
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Probably low	The percentage of missing data for organochlorine pesticides and PCBs was small (3%) and it was adequately addressed using imputation models.
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Probably low	The authors did not report blinding but the study design prevents knowledge of exposure groups.
Can we be confident in the exposure characterization?	Probably low	No analytical information reported in the main text, all the information referred to external publication (Mendez et al. 2011)
Can we be confident in the outcome assessment?	Definitively low	Weight (kilograms) and height (centimeters) of the children at approximately 7 years of age (range, 64–95 months) were measured by specially trained nurses; 470 children participated in this follow-up. Child weight and height were measured using standard protocols (without shoes and in light clothing).
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study's specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	This study was funded by grants from Instituto de Salud Carlos III, Spanish Ministry of Health, Generalitat de Catalunya and RecerCAixa. The authors declare they have no actual or potential competing financial interests.

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S17. Risk of bias of Tang-Peronard et al. 2015, Danish part of the European Youth Heart Study (EYHS) , according instructions reported at NTP/OHAT risk of bias tool (OHAT 2015a). Abbreviations: SMO, smoking; ACT, physical activity; MAT-BMI, maternal body mass index; SEX, gender.

Bias	Risk of Bias	Comments
CONFOUNDING BIAS Did the study design or analysis account for important confounding and modifying variables?	Probably low	The key confounders were adequately addressed or included in the model. Confounding factors included in the model: SMO, ACT, MAT-BMI, SEX
PERFORMANCE BIAS Did researchers adjust or control for other exposures that are anticipated to bias results?	Probably low	Measured levels of PCBs but not included in a multipollutant model. Not from a population of high occupational or acutely contaminating exposures.
ATTRITION/EXCLUSION BIAS Were outcome data incomplete due to attrition or exclusion from analysis?	Probably high	Large number of missing values (n=112)
DETECTION BIAS Were the outcome assessors blinded to study group or exposure level?	Probably low	The authors did not report blinding but the study design prevents knowledge of exposure groups.
Can we be confident in the exposure characterization?	Definitively low	Used gas chromatography using a dual capillary column system with microelectron capture detection which is considered an adequate analytical method. The authors followed quality assurance programs, and the main quality performance parameters were reported in the main text.
Can we be confident in the outcome assessment?	Definitively low	All measurements were carried out using standardized methods
SELECTIVE REPORTING BIAS Were all measured outcomes reported?	Definitively low	All of the study's specified outcomes were adequately reported.
CONFLICT OF INTEREST	Definitively low	This work was supported by the Danish Council for Strategic Research, Program Commission on Health, Food and Welfare. None of the authors had a conflict of interest.

Association between exposure to p,p'-DDT and its metabolite p,p'-DDE and obesity: integrated systematic review and meta-analysis

Table S18. Prospective studies excluded in the meta-analysis assessing associations between the exposure to DDTs and obesity.

Abbreviations: Gen, gender; M, male; F, female; MF, male and female combined, M/F male and female stratified. Cohorts: CPP, Collaborative Perinatal Project; PIVUS, The Prospective Investigation of the Vasculature in Uppsala Seniors; AMICS-INMA-Menorca, Menorca Asthma Multicentre Infants Cohort Study– Infancia y Medio Ambiente. Outcomes: BMI, body mass index; WC, waist circumference; OVE, overweight; OBE, obesity. Statistic: ∅ no statistically significant effect, ▲ statistically significant increase.

Cohort	Ref	Dates	Country	Gen	Age ^a	N	Outcome	Exposure assessment	Exposure levels	Un	Statistic	Exclusion reason
CPP	(Gladden et al. 2004)	61-65	US	M	20	304	BMI	Maternal Serum p,p'-DDE p,p'-DDT	<i>p,p'-DDE quintiles</i> 1.0–2.9 µg/g 3.0–5.9 µg/g 6.0–8.9 µg/g 9.0–11.9 µg/g 12.0–25.1 µg/g	lw	Beta ∅ BMI	Outcome and statistic not combinable in meta-analysis
									* (Cupul-Uicab et al. 2013)	59-65	US	MF/M/F
Michigan fish eaters	(Karmaus et al. 2009)	73 -07	US	F	20-50	176	BMI	Maternal Serum p,p'-DDE	p,p'-DDE <1.5 µg/L 1.503-2.9 µg/L >2.9 µg/L	ww	Beta T2, T3 vs T1 ▲ BMI Monotonic	Outcome and statistic not combinable in meta-analysis
PIVUS	(Lee et al. 2012)	01-09	Sweden	M/F	75	970	OBE WC	Plasma p,p'-DDE	p,p'-DDE 0.011-0.902 µg/L 0.903-1.486 µg/L 1.487-2.304 µg/L 2.305-4.039 µg/L 4.040-23.271 µg/L	ww	OR ∅ OBE WC; ▲ OBE WC males Non-monotonic	Outcome and statistic not combinable in meta-analysis
Faroe Island									(Tang-Peronard et al. 2014)	96-01	Faroe Islands	M/F
AMICS-INMA-Menorca	(Valvi et al. 2012)	97-05	Spain	MF/M/F	6.5	344	OVE	Cord blood p,p'-DDE p,p'-DDT	< 0.7 ng/mL 0.7–1.5 ng/mL > 1.5 ng/mL	ww	RR ▲ T2 Overweight Non-monotonic	Outcome and statistic not combinable in meta-analysis

^a Age at follow-up clinical evaluation (years)

3. NON-HUMAN EVIDENCE

Table S19. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and obesity. Obesity was defined as increased adiposity. Abbreviations: i.p. intraperitoneal injection; o.s. oral administration; PND, post-natal day.

Reference	Species	Strain	Sex	Chemical	Dose levels	Route	Duration/ frequency	Endpoint
(Skinner et al. 2013)	Rats	Hsd:Sprague Dawley	Males and females	p,p'-DDT	25 and 50 mg/kg BW/day	i.p.	Transgenerational Daily, on days 8 to 14 of gestation	Obesity (Body mass increase, adiposity and presence of associated diseases)
(La Merrill et al. 2014)	Mice	C57BL/6J	Males and females	p,p'-DDT* o,p'-DDT	1.7 mg/kg BW/day	o.s.	Prenatal Daily (11.5 day post coitus to PND 5)	Adiposity

* La Merrill et al. 2014; purity neat p,p'-DDT 98.5% and o,p'-DDT ,100%, AccuStandard

Table S20. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and abnormal lipids.

Abbreviations: BW, body weight; CHO, cholesterol; FFA, free fatty acids; GD, gestational day; g.t. gastrointestinal tube; HFD, high fat diet; i.p. intraperitoneal injection; i.t. intratraqueal; NCM, Northern Contaminant Mixture; NEFA, non-esterified fatty acids; NS, no specified; o.s. oral administration; POP, persistent organic pollutants; STD, standard diet; TAG, triacylglycerol.

Reference	Species	Strain	Sex	Chemical	Dose levels	Route	Duration/ frequency	Endpoint
Rats								
(Platt and Cockrill 1967)	Rats	Wistar-derived Alderley Park	Male	DDT	0.1 % w/w diet	g.t.	Daily for 14 days.	Liver enzymatic activities Liver weight
(Platt and Cockrill 1969)	Rats			p,p'-DDT	0.1 % w/w diet	o.s.		Liver weight
(Darsie et al. 1976)	Rats	Sprague-Dawley	Male	p,p'-DDT o,p'-DDT	5 and 100 ppm o,p'-DDT, p,p'-DDT	o.s.	30, 60, or 90 days	Fatty acids
(Appleton et al. 1981)	Rats	Wistar	Male	p,p'-DDT	500 mg/kg	o.s.	14 days	Liver weight Serum cholesterol
(Rao et al. 1981)	Rats	Wistar strain	Male	p,p'-DDT	Acute exposure: 600 mg/kg BW Chronic exposure: 15 mg/kg BW	g.t.	Chronic exposure: 45 days	Serum fatty acids
(Nagaoka et al. 1986)	Rats	Wistar	Male	p,p'-DDT	0.1% DDT	o.s.	7 days	Liver lipids
(Narayan et al. 1990)	Rats	Wistar	Male	p,p'-DDT	0.05 mg/g BW	i.t.	Three consecutive days	Liver lipids
(Okazaki and Katayama 2003)	Rats	NS	Males	p,p'-DDT	0.07 g/100 g of diet	o.s.	2 weeks	Liver and serum lipids Liver enzymes activities
(Okazaki and Katayama 2008)	Rats	Wistar	Male	p,p'-DDT	0.07 g/100 g of diet	o.s.	2 weeks	Liver and serum lipids Liver lipogenic enzymes
(Ishikawa et al. 2015)	Rats	Sprague-Dawley	Male	DDT mixture*	5.6 µg DDT mixture/kg BW/day	o.s.	Once a week for 4 weeks followed by caloric restriction	Fasting plasma TAG, NEFA Postprandial plasma TAG, NEFA
(Rodriguez-Alcala et al. 2015)	Rats	Sprague-Dawley	Male	p,p'-DDE	100 µg/kg BW/day (STD/HFD)	o.s.	12 weeks	Liver fatty acids
Mice								
(La Merrill et al.	Mice	C57BL/6J	Males and	p,p'-DDT*	1.7 mg/kg BW	o.s.	Prenatal	Hepatic lipogenic enzymes

2014)			females	o,p'-DDT			Daily (11.5 day post coitus to PND 5)	Hepatic lipids Conjugated bile acids
(Howell et al. 2015)	Mice	C57BL/6H	Male	p,p'-DDE*	2.0 mg/kg		Daily for five days	Hepatic TAG, CHO Serum TAG, CHO, FFA
Other species								
(Takayama et al. 1999)	Monkeys	Macaca fascicularis and rhesus	Male and female	p,p'-DDT	10 mg/kg		5 days a week, for 130-month	Liver fats
(Benton et al. 1994)	Sailfin molly			o,p'-DDT	1, 10, 25, 50, 75, and 100 pg/liter.		Once a day, 21 days	Lipid content
(Westlake et al. 1979)	Japanese quail		Female	DDMU	20, 100, 350 and 1000 mg/kg BW			Lipid analysis

* Ishikawa et al. 2015; purity neat p,p'-DDT 97.4%, o,p'-DDT , 99.7% and p,p'-DDE, 100%, AccuStandard. Howell et al. 2014; purity p,p'-DDE 98%; Chem Service. La Merrill et al. 2014; purity neat p,p'-DDT 98.5% and o,p'-DDT ,100%, AccuStandard

Table S21. Summary of study characteristics of in vivo evidence reporting associations of exposure to DDTs and energy balance and adipokines. Abbreviations: BW, body weight; g.t., gastrointestinal tube ; i.p., intraperitoneal; i.v., intravenous; o.s., oral administration.

Reference	Species	Strain	Sex	Chemical	Dose levels	Route	Duration/ frequency	Endpoint
Adipokines								
(Howell et al. 2014)	Mice	C57BL/6H	Male	p,p'-DDE*	0.4 and 2.0 mg/kg	g.t.	Daily for five days	Leptin, resistin
(Howell et al. 2015)	Mice	C57BL/6H	Male	p,p'-DDE*	2.0 mg/kg	g.t.	Daily for five days	Leptin, resistin and adiponectin
Energy balance								
(Ahdaya et al. 1976)	Mice		Female	DDT	LD ₂₅ and LD ₅₀	i.p.	Single dose	Rectal temperature
(Ishikawa et al. 2015)	Rats	Sprague–Dawley	Male	DDT mixture*	5.6 µg DDT mixture/kg BW/day	o.s.	Once a week for 4 weeks followed by caloric restriction	Core body temperature
(La Merrill et al. 2014)	Mice	C57BL/6J	Males and females	p,p'-DDT* o,p'-DDT	1.7 mg/kg BW	o.s.	Prenatal Daily (11.5 day post coitus to PND 5)	Energy expenditure

* Ishikawa et al. 2015; purity neat p,p'-DDT 97.4%, o,p'-DDT , 99.7% and p,p'-DDE, 100%, AccuStandard. Howell et al. 2014; purity p,p'-DDE 98%; Chem Service. La Merrill et al. 2014; purity neat p,p'-DDT 98.5% and o,p'-DDT ,100%, AccuStandard

Table S22. Relationship of parts per million in diet to mg/kg body weight per day (JECFA 2000).

Animal	Weight (kg)	Food consumption (g/day)	Type of diet	1 mg/kg BW/day is equivalent to, in ppm of the diet	1 ppm in food is equivalent to, in mg/kg BW per day
Mouse	0.02	3	Dry	7	0.150
Rat, young	0.10	10	laboratory	10	0.100
Rat, older	0.40	20	chow diets	20	0.150

Table S23. Summary of study characteristics of in vitro evidence reporting associations of exposure to DDTs and adipogenesis, lipogenesis and markers of metabolic homeostasis. Abbreviations: ATGL, adipose triglyceride lipase; DMSO, dimethyl sulfoxide, PPAR, peroxisome proliferator-activated receptor; CEBP enhancer-binding protein; Lep, leptin; LpL, lipoprotein lipase; Insig1, Insulin-induced gene-1; Fapb4, Fatty acid binding protein 4; Fasn, fatty acid synthase; Srebf1, sterol regulatory element-binding protein 1c; Slc2a4, glucose transporter type 4.

Author	Cell line, cell type, or tissue	Chemical	Concentration levels	Vehicle	Endpoint
(Chapados et al. 2011)	Human subcutaneous preadipocytes	p,p'-DDE	5, 50 and 500 μ M	DMSO	Proliferation
(Moreno-Aliaga and Matsumura 2002)	Adipocytes 3T3-L1 and 3T3-F442	p,p'-DDT	0, 1, 10, 20, 30, 50 μ M	Ethanol	Adipogenic differentiation Expression of adipogenic proteins PPAR γ -1 and PPAR γ -2 Binding activity of the C/EBP proteins (α , β , and δ)
(Strong et al. 2015)	Isolated mesenchymal stem cells (MSCs) MCF7 cells	p,p'-DDT	100 pM, 1 nM, 10 nM, 100 nM, 1 μ M, or 10 μ M.	DMSO	Cell viability Osteogenic and adipogenic differentiation qRT-PCR GLUT4, Leptin, LpL, PPAR γ Gene expression and hierarchical clustering analysis.
(Howell and Mangum 2011)	Adipocytes NIH3T3-L1	p,p'-DDE	2 or 20 μ M	DMSO	Adipogenesis assay Fatty acid uptake Lipolysis assay Cytokine/adipokine multiplex immunoassay Leptin, resistin, and adiponectin expression
(Ibrahim et al. 2011)	Isolated soleus muscle Adipocytes 3T3-L1	p,p'-DDE	10 nM, 1 μ M	NS	Lipid accumulation was quantified using Oil Red O staining.
(Taxvig et al. 2012)	Adipocytes 3T3-L1 Adipocytes NIH-3T3	p,p'-DDE	0.3, 1, 3, 10, 30, 100 μ M	DMSO	Adipocyte differentiation assay PPAR transactivation Adipokine release (leptin, resistin, adiponectin)
(Mangum et al. 2015)	Adipocytes 3T3-L1	p,p'-DDE	0, 3, 10, 30, 100 μ M	DMSO	Adipocyte differentiation Gene expression PPAR γ , REBP1C, FASN, FABP4, LEP

Table S24. Summary table of risk of bias of in vivo studies. Classification: (++) definitively high risk of bias, (+) probably high risk of bias, (-) probably low risk of bias, (--) definitively low risk of bias. (*) Asterisk indicates a key risk of bias domain. T1 means tier 1 and T2, tier 2 according the NTP/OHAT (2015) tiered approach risk of bias tool approach. Risk of bias ratings according Navigation Guide instructions for non-human studies (Kousta et al. 2014), full instructions at “Section 7, Instructions to assess the risk of bias of in vivo studies”.

	(La Merrill et al. 2014)	(Skinner et al. 2013)	(Okazaki and Katayama 2003)	(Okazaki and Katayama 2008)	(Ishikawa et al. 2015)	(Rodriguez-Alcala 2015)	(Howell et al. 2014)	(Howell et al. 2015)
Sequence generation* Was the allocation sequence adequately generated?	--	+	+	+	-	-	+	--
Allocation concealment Was allocation adequately concealed?	+	+	+	+	+	+	+	+
Blinding of personnel and outcome assessors* Was knowledge of allocated interventions adequately prevented?	--	--	+	+	+	+	+	+
Incomplete outcome data Were incomplete outcome data adequately addressed?	--	--	--	-	-	-	-	-
Selective outcome reporting Were study reports free of selective outcome reporting?	-	-	--	-	-	-	-	-
Other potential threats to validity Was study free of other problems regarding risk of bias?	--	--	-	--	--	--	--	--
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	--	--	+	-	-	--	--	--
Summary Tiered Classification	T1	T2	T2	T2	T2	T2	T2	T2

Table S25. Risk of bias of the study La Merrill et al. 2014. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Low risk	<i>“we randomized mice into 2 study arms” “For the high fat feeding study, littermates were randomized to high fat diet...” “GTT...randomized selection from n =15 DDT- and n= 14 vehicle-exposed litters...”</i>
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Low risk	Histopathological evaluations were assessed by a veterinary pathologist who was blinded to the treatments.
Incomplete outcome data Were incomplete outcome data adequately addressed?	Low risk	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	Methods to control litter-effects were reported (Mixed linear models)
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Low risk	The authors have declared that no competing interests exist. This research was supported through funding by the National Institute of Health, the American Diabetes Association, and United States Department of Agriculture

Table S26. Risk of bias of the study Skinner et al. 2013. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably high	Randomization not discussed
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Low risk	All the histology was blinded.
Incomplete outcome data Were incomplete outcome data adequately addressed?	Low risk	No missing data reported directly by the author. All the data was uploaded to NCBI.
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	Methods to control litter-effects were reported (“Individual animals from different litters were used for analysis”)
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Low risk	The authors declare they have no competing interests. This study was supported by a grant from the NIH, NIEHS to MKS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Table S27. Risk of bias of the study Okazaki and Katayama 2003. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably high	Randomization not discussed
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Probably high	The authors did not provide a conflict of interest declaration statement, or funding sources.

Table S28. Risk of bias of the study Okazaki and Katayama 2008. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably high	Randomization not discussed
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Probably low	This work was supported by a Sasakawa scientific research grant from The Japan Science Society (grant no. 16-334).

Table S29. Risk of bias of the study Ishikawa et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably low	<i>“obese SD rats were randomized into two treatment groups (control or DDT mixture) matched for body weight and fasting lipid concentrations”</i>
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Probably low	This research was supported by NIH (ES019919 to ML; DK095980,HL091333, HL107256, HL107256 to PH), a University of California,Davis College of Agricultural and Environmental Sciences Program-matic Initiative), and a Multi-campus grant from theUniversity of California, Office of the President (#142691 to PH).

Table S30. Risk of bias of the study Rodriguez-Alcala et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably low	<i>“randomly divided into four groups of six animal”</i>
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Low risk	The authors declare no competing financial interest.

Table S31. Risk of bias of Howell et al. 2014. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Probably high	Randomization not discussed
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Low risk	The authors declare no competing financial interest. The work was funded by the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) under award number R15ES019742.

Table S32. Risk of bias of Howell et al. 2015. Risk of bias rating according Navigation Guide instructions for non-human studies (Kousta et al. 2014).

Bias domain	Rate	Comment
Sequence generation Was the allocation sequence adequately generated?	Low risk	<i>animals were randomly divided into one of four experimental groups (n = 15/group)</i>
Allocation concealment Was allocation adequately concealed?	Probably high	Not reported
Blinding of personnel and outcome assessors Was knowledge of allocated interventions adequately prevented?	Probably high	Blinding not discussed
Incomplete outcome data Were incomplete outcome data adequately addressed?	Probably low	Allocation numbers pre-specified in methods section and adequately followed
Selective outcome reporting Were study reports free of selective outcome reporting?	Probably low	Outcomes outlined in abstract/methods section of paper were reported in results section
Other potential threats to validity Was study free of other problems regarding risk of bias?	Low risk	No other potential biases are suspected.
Conflict of interest Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?	Low risk	There are no conflicts of interest by any authors. The work was funded by the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) under award number R15ES019742.

4. RATING AND INTEGRATION OF EVIDENCE

4.1. INSTRUCTIONS FOR RATING THE CONFIDENCE

We have applied the NTP/OHAT framework, based on the GRADE guidelines, to rate the confidence with the body of evidence, translate to a level of evidence and integrate the different streams of evidence to deliver the hazard identification conclusions (OHAT 2015b). The overall work-flow process is illustrated in the Figure 1 of the main manuscript, considering two main bodies of evidence (human and *in vivo* studies) addressing the main health outcomes and we considered a supplemental body of evidence with mechanistic data from *in vivo* and *in vitro* studies reporting mechanistic events and secondary outcomes related with obesity. The quality and level of evidence was evaluated independently for human and animal evidence, establishing an initial confidence rate and using a sequential process considering those factors that may affect (upgrading or downgrading) the confidence including the risk of bias, imprecision, publication bias, indirectness, magnitude, dose-response and plausible confounding.

The risk of bias was evaluated by means of risk of bias tools specifically designed for human epidemiological studies and animal studies and adapted for DDTs and obesity outcomes (Kousta et al. 2014; OHAT 2015b). We did not assess the risk of bias of *in vitro* studies due to the lack of risk of bias tools or guidance to assess the internal quality; however we considered the other factors to rate the confidence (Rooney et al. 2016).

The final confidence rating of each body of evidence (human and *in vitro*) was translated to a level evidence, considering the direction of the effect (“health effect” or “no health effect”), and integrated using the hazard identification scheme to provide a preliminary classification of the chemical (“known”, “presumed”, “suspected” or “not classifiable” hazard for humans).

Two supplemental bodies of evidence were established with supporting *in vivo* studies (reporting secondary outcomes) and *in vitro* studies and rated similarly to the main bodies of evidence to establish a final level of evidence. We considered that a high level of supporting evidence may upgrade the preliminary classification, while a low level of evidence could downgrade. Moderate level of evidence does not modify the rate. We judged together both bodies of supporting evidence integrating both levels of evidence to deliver a final decision to upgrade or downgrade the final rate. The details on the confidence rating process is described in detail in the protocol at Section1.

4.2. RATING THE CONFIDENCE IN THE OF BODY OF EVIDENCE FROM HUMAN STUDIES AND LEVEL OF EVIDENCE FOR HEALTH EFFECT

4.2.1. INITIAL RATING OF CONFIDENCE

Observational prospective studies meet the three following features, therefore initially classified as “moderate” confidence, according the NTP/OHAT classification setting (OHAT 2015b).

- The exposure assessment demonstrates that exposures occurred prior to the development of the outcome (or concurrent with aggravation/amplification of an existing condition).
- The outcome is assessed on the individual level (i.e., not through population aggregate data).
- An appropriate comparison group is included in the study.

4.2.2. POTENTIAL FACTORS THAT MAY REDUCE CONFIDENCE

Risk of bias

The main source identified to increase the risk of bias was the performance bias due to the extended use of single-pollutant models when simultaneous exposure to complex mixtures of obesogenic compounds is highly suspected or even reported. The methodological limitations to address collinearity and multicollinearity, measurement errors, pollutants interactions and potential non-linear exposure-health relationships, may be some of the determinant factors to bring the authors choosing the single-pollutant approach as the preferred model (Billionnet et al. 2012). We overall considered that most of studies may be at probably high risk of performance bias with only three exceptions that used multipollutant models. Based on a preliminary search on the literature we elaborated a directed acyclic graph (DAG) to identify relevant covariates (Figure S2), and we selected as potential key confounders the maternal BMI, maternal smoking and sex which were controlled by all studies.

The studies retained for meta-analysis addressed potential confounding bias by adjusting for known confounders in multivariate regression models (Table S8). Most studies adjusted for maternal BMI, or occasionally for maternal weight and/or height. Most analyses also included adjustment for maternal- age, education, parity, breastfeeding, and an indicator of socioeconomic status (race, education, income, social class, and/or socioeconomic index). Birth weight was also included in the model of two studies (Agay-Shay et al. 2015; Vafeiadi et al. 2015). Physical activity and/or diet were adjusted in models of three studies (Agay-Shay et al. 2015; Hoyer et al. 2014; Tang-Peronard et al. 2015). Maternal smoking was modeled as a confounder in the majority of studies retained here (Cupul-Uicab et al. 2010; Delvaux et al. 2014; Hoyer et al.

2014; Vafeiadi et al. 2015) with the exception of one study in which maternal smoking did not modify the effect estimate (Warner et al. 2014). One study concluded that risk of obesity associated with DDTs would be exacerbated by maternal smoking (Cupul-Uicab et al. 2010). Maternal alcohol consumption was included as a confounder in the regression models of one study (Hoyer et al. 2014). Additionally, we considered the lack of control over some important covariates commonly related with energy balance and weight gain, such as diet and exercise. Some studies like Warner et al. 2015 considered a large list of variables including diet and exercise and finally they did not include those variables in the model because the lack of contribution to the model variance (common criteria of exclusion <10%).

Attrition/exclusion bias was not suspected because the studies presented commonly small missing data and preventing bias by managing the censored datasets adequately. Detection bias was not suspected due the studies reported strategies to blind assessors and the exposure and outcomes were assessed using robust and validated methodologies. Selective reporting bias was judged to be unlikely as all outcomes measured were reported.

We overall classified all the studies in the first tier of risk of bias because we did not suspect of risk of bias in the key domains (definitively/probably low risk of confounding and detection bias) and most of other domains classified also as definitively or probably low risk of bias, with only the exception of performance bias (probably high risk of bias), further details in the Tables S10-17. The overall risk of bias was considered to be “Not likely”.

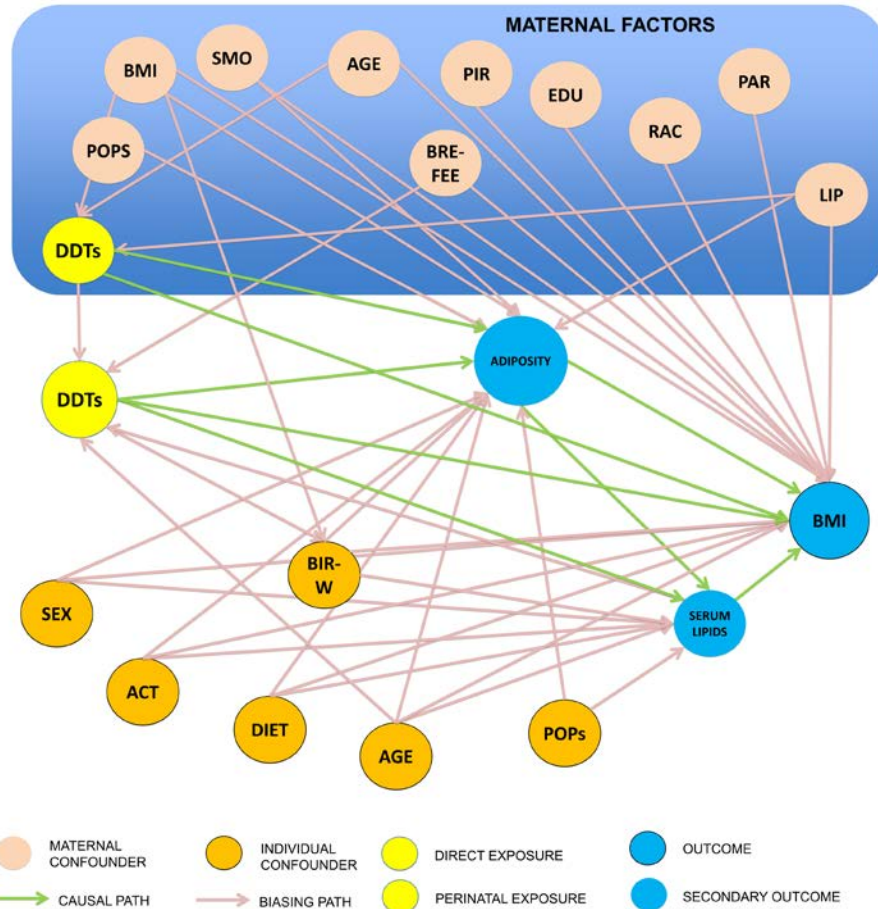


Figure S7. Directed acyclic graph illustrating the causal associations between different exposures (direct and maternal), covariables and outcomes in the model where the levels of DDTs are adjusted by lipids (lipid weight basis). Covariables: ACT, physical activity; BIR-W, birth weight; BRE-FEE, breastfeeding; EDU, education; LIP, lipids; AGE, age; BMI, body mass index; PAR, parity; POPs: other persistent organic pollutants; RAC, race; SMO, smoking.

Unexplained inconsistency

The between studies- low heterogeneity (I^2 39.5%) and variance ($\tau^2 < 0.013$) were not considered concerning to downgrade the confidence by unexplained inconsistency.

Directness/applicability

We did not penalize the confidence rating on this regards because the human studies assessed target human population, health outcomes and exposures of interest. Despite some differences between exposure windows, ages or exposure assessment approaches, the stratification analysis did not show differences to modify our conclusions.

Imprecision

Judgement of imprecision was discussed based on the assumptions done to combine studies in the meta-analysis with some discrepant methodological approaches. The limited number of available studies with identical methodological approaches constrained to make assumptions, decreasing accuracy of estimates in benefit of having a more representative sample of population, preventing the loss of information resulting from a biased sample of studies. Sources of variability between studies included the population characteristics, matrices, lipid-normalization of exposure estimates or covariates included in the multivariate regression model. Combining regression coefficients from different exposure estimates (lipid-normalized or expressed in wet weight) was one of the critical issues that could impair the accuracy of the final meta-estimates. To overcome the effect of pooling those effect sizes we further stratified the studies comparing the resulting meta-estimates, and the results showed slightly higher meta-estimates for the strata of studies using wet weight exposure but in the similar range (confidence intervals were commonly overlapping). The number of studies is considered concerning to cause imprecision for continuous variables when it goes below 400 participants (Kulig et al. 2012). The number of participants in the included studies averaged ~450 so we considered that the imprecision was not serious because only few studies were considered underpowered and providing larger confidence intervals. When those studies were omitted in the sensitivity analyses, the conclusions did not change substantially.

Publication bias

The funnel plots did not show asymmetry and the Egger's test was not significant. Also, considering the absence of private funding or conflict of interest, as well as, the lack of potentially unpublished studies (e.g. conference abstracts, grey literature), we determined publication bias was not serious.

4.2.3. POTENTIAL FACTORS THAT MAY INCREASE CONFIDENCE

Magnitude

We concluded that the magnitude of the effect was modest and thus did not justify upgrading the confidence rating. GRADE approach classifies the large magnitude of that effect when the relative risk ranges are between 2 and 5, and very large when is higher than 5.

Dose-Response

After a preliminary examination of the dose-response trend among individual studies, an inverted U-shaped dose-response curve was observed in some studies. However, a consistent trend was not exhibited across individual studies. We did not perform dose-response meta-analysis due to the variability between the different exposure groups and reference groups across the different studies.

Residual confounding

The GRADE and NTP/OHAT approaches upgrade the confidence when the study reports an effect or association despite the presence of residual confounding (OHAT 2015b). We were especially concerned about the potential over-adjustment bias resulting from normalizing the exposure levels of DDTs by lipids. We included in the meta-analysis 14 studies providing the exposure assessment levels in lipid basis, however this approach has been demonstrated to provide more biased results than those models using wet values and including the lipid concentration as a covariate in the model (Gaskins and Schisterman 2009; Schisterman et al. 2005). The stratification analysis also suggested that the meta-estimates from the studies in wet weight could be higher, despite the lower statistical potency that could attenuate such results. Serum lipids are in the same causal pathway of adiposity and body mass index, and associated with serum exposure levels (Patel et al. 2012), as illustrated in the DAG for the proposed model considering other covariates (Figure S1). While the lipid adjusted concentration of serum levels is an extensively implemented approach for lipophilic compounds, the bias towards the null is suspected in those cases when the intermediate variable is in the same causal pathway (Schisterman et al. 2009). Additionally, the lipid standardization is subject of several assumptions subject of criticism. The standardization by the lipid levels assume a state of equilibrium and is frequently performed dividing the serum levels of DDTs by the serum total lipid levels assuming linear correlations (Phillips et al. 1989; Porta et al. 2009). However, neither the state of body equilibrium and linear correlations may be expected in the scenario where DDTs exhibit strong disrupting effects of lipid homeostasis. The results from *in vivo* studies also supported the potential effect on lipid homeostasis, being highly consistent in liver, however the results were not so consistent among circulating lipid levels. We judged the bias is likely to occur however, the evidence to support the upgrading is scarce.

Consistency across populations

Despite the results were consistent among both species, and in turn with the human epidemiological results, we did not judge to upgrade because the limited number of available studies to conclude such relationship.

4.2.4. FINAL RATING OF CONFIDENCE

After balancing the upgrading and downgrading factors, the final rating of the confidence with the body of evidence was finally appraised to be “moderate”.

4.2.5. LEVEL OF EVIDENCE FOR HEALTH EFFECT FROM HUMAN STUDIES

We used the OHAT descriptors to establish the translation the “moderate” confidence of the body of evidence to “moderate” level of evidence of health effect, between the exposure to p,p'-DDE and obesity considering the direction of the effect to the presence of “health effect”.

4.3. RATING THE CONFIDENCE IN THE BODY OF EVIDENCE FROM PRIMARY *IN VIVO* STUDIES AND LEVEL OF EVIDENCE FOR HEALTH EFFECT

4.3.1. INITIAL RATING OF CONFIDENCE

The initial quality level of experimental animal data will be considered as “high”, comparable to human randomized controlled trials.

4.3.2. POTENTIAL FACTORS THAT MAY REDUCE CONFIDENCE

Risk of bias

Both studies were rated at low or probably low risk of bias for most of bias domains, which included their proper considerations of litter effects. The exemption was that one study was classified as probably high risk of bias in the sequence generation domain due to the lack of randomization of treatments (Table S24). Overall, we judged the risk of bias of this body of evidence to be “serious” because one study was classified in the Tier 1 and the other in the Tier 2.

Unexplained inconsistency

The results from both experimental studies had some relevant inconsistencies. For instance, Skinner et al. (2013) observed obesity only in the third and fourth generations, whereas La Merrill et al. (2014) reported increased adiposity in the first generation. Inconsistencies in the methodological approaches (e.g. timing, dose and route of exposure, rodent model) may explain these disparities however because there are only two studies we conclude consistency is unknown in accordance with OHAT guidance (2015a), thus we did not downgrade due to inconsistency.

Directness/applicability

We considered the directness and applicability of the animal model in humans and the concentration doses (Figure S2). In case of La Merrill et al. (2014) the 1.7 mg DDT/kg body weight resulted in the internal concentration level of 2.2 ± 0.1 ng p,p'-DDE/mL which is in the middle range of exposure of the human epidemiological studies. Accurate assessment of human relevance of exposure doses used by Skinner et al. (2013) entails a complex exercise because the lack of monitoring of internal dose across the three generations. However, we approached roughly to the exposure at F₂ (only F₃ showed positive associations) from the concentration doses in F₀ of 25 and 50 mg/kg BW/day. We applied the rates of transplacental and lactational transfer of p,p'-DDE in Sprague-Dawley rats reported by You et al. (1999) and we estimated that the internal exposure at F₁ may expected in the high range of the human cohorts and F₂ and F₃

in the middle and lowest range of exposure (You et al. 1999). Both studies used rodent models (C57BL/6J mice and Hsd:Sprague Dawley rats) which are considered as of direct applicability for human health. According GRADE guidelines, downgrading by indirectness may be only justified when there is some compelling reason to suspect the different biology could modify the magnitude of the effect, thus we rated by zero (Guyatt et al. 2011e).

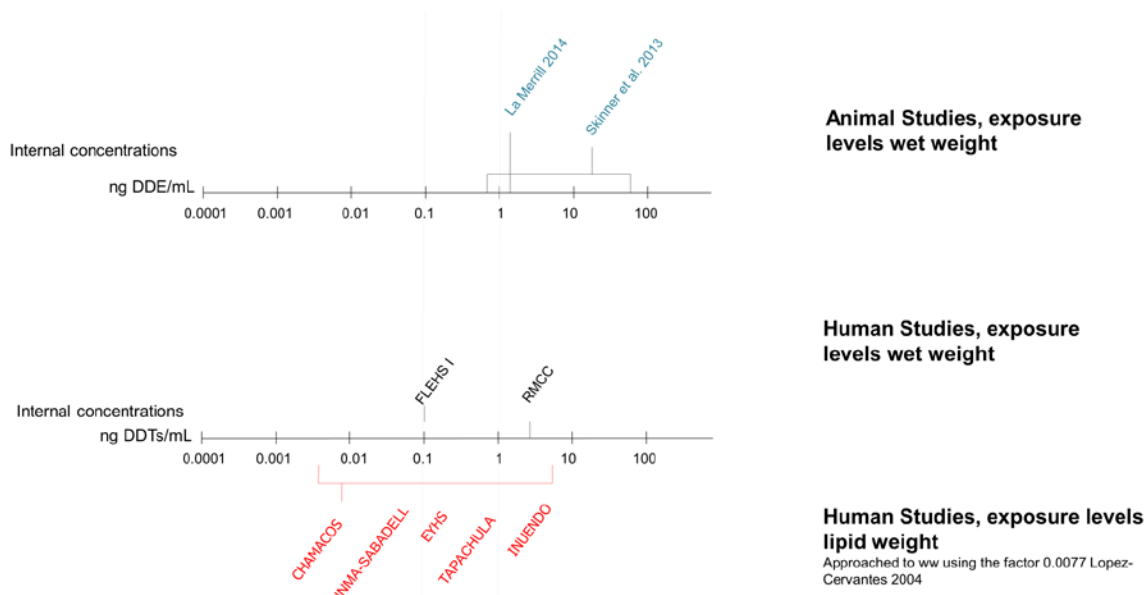


Figure S8. Summary of exposure levels from in vivo studies compared with human epidemiological studies to assess directness of levels. The internal doses of human studies are serum levels expressed in wet weight (black) and converted to wet weight from reported levels in lipid weight (red) using the conversion factor 1:129.8 wet weight:lipid weight (Lopez-Carrillo et al. 1999). Two approaches were assumed to compare the exposure levels of *in vitro* studies: level of cell culture exposure assuming accumulation of p,p'-DDT and p,p'-DDE in adipose tissues and using a ratio 1:129.8 serum:adipose tissue; and assuming the exposure in adipocytes without accumulation using a ratio 1:1 serum:adipose tissue.

Imprecision

An acceptable number of animals per treatment and controls were used in both studies (n=15, La Merrill et al., 2014; n=30, Skinner et al., 2013), providing accurate estimates with narrow error bars; accordingly, we decided imprecision was not serious.

Publication bias

We judged no reason to suspect of publication bias. The results were consistent throughout the years and regardless of size. We did not suspect of unpublished studies considering the results from the comprehensive literature search, including conference abstracts and grey literature. The studies were funded by governmental and/or other public sources, and conflicts of interest were not stated.

4.3.3. POTENTIAL FACTORS THAT MAY INCREASE CONFIDENCE

Magnitude

Considering the rates of obesity from Skinner et al. (2013), gave us a crude risk ratio (95% CI) of 14.5 (2.1-102.6) for females and 3.4 (1.7-6.7) among males. The magnitude of the effect on La Merrill et al (2014) was only possible to evaluate by means of the p-values resulting from the statistical analysis which was estimated at <0.05, revealing a modest magnitude. Despite the large magnitude of the effect reported by Skinner et al. (2013), the large confidence interval computed for males, revealed relevant imprecision of such risk estimate. According the GRADE guidelines, associations (RR, relative risk) greater than 2 would justify the increase of one category, and greater than 5 up to two categories (Grade 9), however and considering the disparity of these results we concluded to do not upgrade.

Dose-Response

La Merrill et al. (2014) tested a single dose level of 1.7 mg/kg bw administered prenatally and compared to the control, thus dose-response was applicable. In case of Skinner et al. (2013) two dose levels (25 and 50 mg/kg bw) were tested prenatally, and followed-up for three generations (F₀ to F₃). The increase of obesity rates compared to the control was only apparent at F₃ following a monotonic trend among males and non-monotonic trend among females. We concluded the evidence is limited to conclude a consistent dose-response trend for upgrading.

Residual confounding

We did not identify sources of residual confounding that may justify upgrading the confidence.

Consistency across species/models

The results were consistent among both species, and in turn with the human epidemiological results, we did not judge to upgrade because the limited number of studies to conclude such relationship.

4.3.4. FINAL RATING OF CONFIDENCE

Overall, we considered the main body of evidence from animal studies, assembled by two studies, was biased and downgraded the preliminary classification from “high” to “moderate” confidence.

4.3.5. LEVEL OF EVIDENCE FOR HEALTH EFFECT FROM *IN VIVO* STUDIES

The nature or direction of the effect was to a “health effect”, thus the confidence was translated to a “moderate” level of evidence of obesogenic effects of DDTs in in vivo studies

4.4. RATING THE CONFIDENCE IN THE BODY OF SUPPORTING EVIDENCE FROM *IN VIVO* STUDIES

4.4.1. INITIAL RATING OF CONFIDENCE

We considered a sort of *in vivo* studies reporting secondary health outcomes of obesity to support the biological plausibility [Table S20 and S21]. The secondary outcomes include abnormal lipids in blood and liver, adipokines and energy balance. We established a preliminary rate of “high” confidence with the body of evidence based on the features of animal study design.

4.4.2. FACTORS AFFECTING THE CONFIDENCE

Unexplained inconsistency

The results from the different studies evaluating the same effect were consistent on the direction and magnitude. Some inconsistent results could be mainly explained by differences on study design, animal models or exposures.

Directness/applicability

We considered two factors related with the directness of supplemental *in vivo* studies: applicability of secondary outcomes and relevance of dose ranges. We judged that even the use of secondary outcomes such as abnormal lipids or adipokines, may not be reflecting accurately the magnitude and causal direction of the effect, the energy balance has a central role in the obesity etiology. The applicability of the concentration ranges was discussed since a relevant sort of studies was performed using high doses of DDTs. However, other studies performed at lower doses depicted similar patterns, thus we overall judged to do not downgrade (Table S1).

Table S33. Summary of dose levels as reported by the supporting *in vivo* studies.

Reference	Species	Chemical	Dose levels
(Ahdaya et al. 1976)	Mice	DDT	LD ₂₅ and LD ₅₀
(Appleton et al. 1981)	Rats	p,p'-DDT	500 mg/kg
(Benton et al. 1994)	Sailfin molly	o,p'-DDT	1, 10, 25, 50, 75, and 100 pg/liter.
(Darsie et al. 1976)	Rats	p,p'-DDT o,p'-DDT	5 and 100 ppm o,p'-DDT, p,p'-DDT
(Howell et al. 2014)	Mice	p,p'-DDE	0.4 and 2.0 mg/kg
(Howell et al. 2015)	Mice	p,p'-DDE	2.0 mg/kg
(Ishikawa et al. 2015)	Rats	DDT mixture	5.6 µg DDT mixture/kg BW/day
(La Merrill et al. 2014)	Mice	p,p'-DDT o,p'-DDT	1.7 mg/kg BW
(Nagaoka et al. 1986)	Rats	p,p'-DDT	0.1% DDT
(Narayan et al. 1990)	Rats	p,p'-DDT	0.05 mg/g BW
(Okazaki and Katayama 2003)	Rats	p,p'-DDT	0.07 g/100 g of diet
(Okazaki and Katayama 2008)	Rats	p,p'-DDT	0.07 g/100 g of diet
(Platt and Cockrill 1967)	Rats	DDT	0.1 %w/w diet

(Platt and Cockrill 1969)	Rats	p,p'-DDT	0.1 % w/w diet
(Rao et al. 1981)	Rats	p,p'-DDT	Acute exposure: 600 mg/kg BW Chronic exposure: 15 mg/kg BW
(Rodriguez-Alcala et al. 2015)	Rats	p,p'-DDE	100 µg/kg BW/day (STD/HFD)
(Takayama et al. 1999)	Monkeys	p,p'-DDT	10 mg/kg
(Westlake et al. 1979)	Japanese quail	DDMU	20, 100, 350 and 1000 mg/kg BW

Risk of bias/ internal validity

Considering most studies have “probably high” risk of bias for sequence generation, blinding and allocation concealment, the studies were classified in the Tier 2 and we downgraded the confidence considering the overall risk of bias to be “Serious”.

Publication bias

We judged no reason to suspect of publication bias. The results were consistent throughout the years and regardless of size. We did not suspect of unpublished studies considering the results from the comprehensive literature search, including conference abstracts and grey literature. The studies were funded by governmental and/or other public sources, and conflicts of interest were not stated.

Magnitude of effect

The magnitude of effect was modest in most studies and no reason to upgrade the confidence.

Dose-response

Few studies assessed the dose-response including concentration points for the endpoints tested.

Consistency

The available evidence for the increased abnormal lipids in liver and impaired thermogenesis by DDTs was consistent across animal species and with the meta-analysis in human studies. However, the lack of consistency on serum lipids disruption and absence of effects on adipokines levels prevented us from upgrading the confidence.

4.4.3. FINAL RATING OF CONFIDENCE

Overall, we considered the supporting body of evidence from animal studies, was biased by the risk of bias, thus downgraded from high to moderate confidence.

4.4.4. LEVEL OF EVIDENCE FOR HEALTH EFFECT FROM *IN VIVO* STUDIES

The nature or direction of the effect was to a “health effect”, thus the confidence was translated to a “moderate” level of evidence of obesogenic effects of DDTs in supporting in vivo studies

4.5. RATING THE CONFIDENCE IN THE BODY OF SUPPORTING EVIDENCE FROM *IN VITRO* STUDIES

4.5.1. INITIAL RATING OF CONFIDENCE

We classified the body of evidence with an initial level of high confidence, and subsequently we assessed the different modifying factors. The final rate was not modified by the factors and thus the high confidence rating was translated into a high level of confidence.

4.5.2. FACTORS AFFECTING THE CONFIDENCE

Unexplained inconsistency

Among downgrading and upgrading factors, we noted a lack of consistency among the results of adipogenic differentiation caused by p,p'-DDE. Only half of the results showed statistically significant increases and the positive results were not consistent across overlapping dosing concentrations (Ibrahim et al. 2011; Mangum et al. 2015). Similarly, lack of consistency extended to the effects of p,p'-DDE on mRNA expression of the main master regulator of adipogenic differentiation PPAR γ (Figure 5B). Despite that the differentiation and *Pparg* expression results had a generally consistent increase with p,p'-DDT exposure, we decided to downgrade due to inconsistency in p,p'-DDE given risk of bias could not be assessed here but was deemed serious in all other experimental streams of evidence evaluated.

Directness/applicability

Most of *in vitro* studies included assessed the effect of p,p'-DDE on adipogenic differentiation and/or markers of lipid metabolism. Both endpoints are directly related with the pathophysiology obesity and pathways related.

The doses tested in the cell culture model in the range of nM to μ M, being the positive results given by the wide range of exposure (0.01 to 100 μ M). Despite, it is difficult to establish the accurate level of internal exposure of adipocytes, the lower and mid-range of exposure are likely to be of biological relevance for human (Figure S3). We overall judged to do not downgrade.

Risk of bias/ internal validity

We did not assess the risk of bias of *in vitro* studies because the lack of standardized guidance applicable to this stream of evidence.

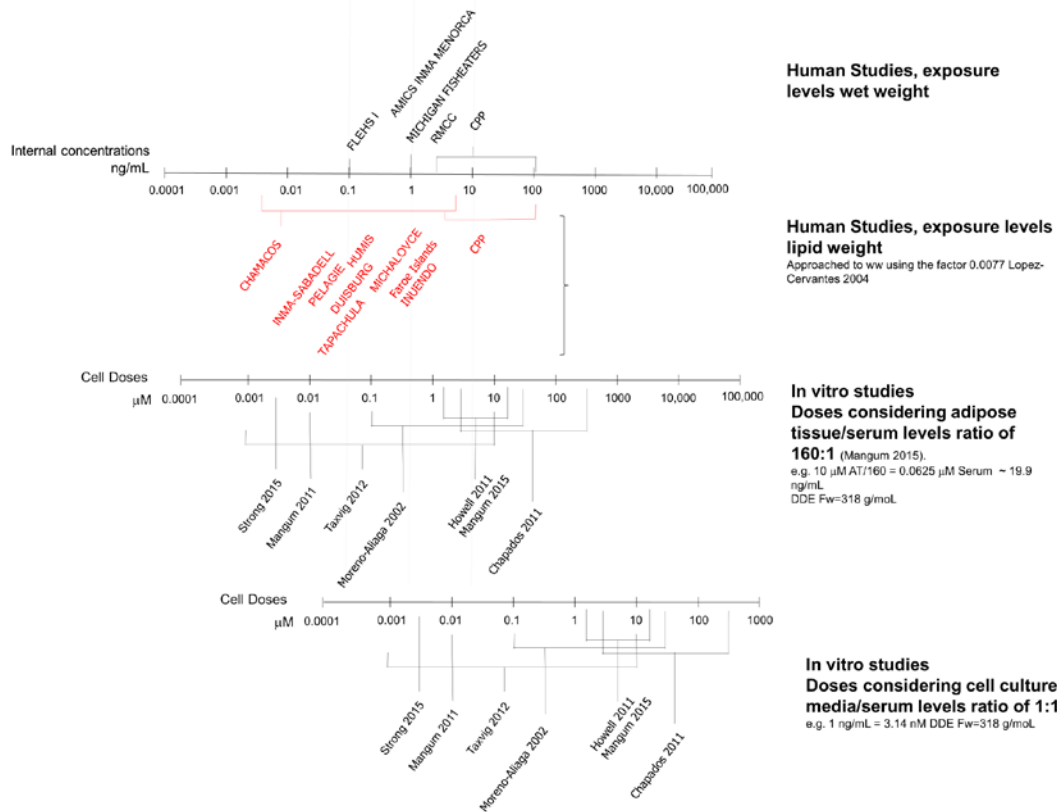


Figure S9. Summary of exposure levels from in vitro studies compared with human epidemiological studies to assess directness of levels. The internal doses of human studies are serum levels expressed in wet weight (black) and converted to wet weight from reported levels in lipid weight (red) using the conversion factor 1:129.8 wet weight:lipid weight (Lopez-Carrillo et al. 1999). Two approaches were assumed to compare the exposure levels of *in vitro* studies: level of cell culture exposure assuming accumulation of p,p'-DDT and p,p'-DDE in adipose tissues and using a ratio 1:129.8 serum:adipose tissue; and assuming the exposure in adipocytes without accumulation using a ratio 1:1 serum:adipose tissue.

Publication bias

We judged no reason to suspect of publication bias. The results were consistent throughout the years and regardless of size. We did not suspect of unpublished studies considering the results from the comprehensive literature search, including conference abstracts and grey literature. The studies were funded by governmental and/or other public sources, and conflicts of interest were not stated.

Magnitude of effect

The magnitude of effect was modest in most studies.

Dose-response

Monotonic dose-response was reported by some studies but there was not a clear trend to justify upgrading.

Consistency

Some unexplained inconsistencies explained in the section “4.2.1.” justified the considerations to downgrade the confidence.

4.5.3. FINAL RATING OF CONFIDENCE

Overall, we considered the supporting body of evidence from *in vitro* studies, was biased by the unexplained inconsistency, thus downgraded from high to moderate confidence.

4.5.4. LEVEL OF EVIDENCE FOR HEALTH EFFECT FROM *IN VIVO* STUDIES

The nature or direction of the effect was to a “health effect”, thus the confidence was translated to a “moderate” level of evidence of obesogenic effects of DDTs in supporting *in vitro* studies

4.6. INTEGRATION OF HUMAN, *IN VIVO* AND *IN VITRO* EVIDENCE AND HAZARD IDENTIFICATION CONCLUSIONS

Considering that both streams of main evidence (human and *in vivo* studies) provided a “moderate” level of evidence, respectively, we established a preliminary classification of p,p'-DDT and p,p'-DDE as “presumed” to be obesogenic for humans.

We considered a moderate level of evidence from supporting *in vivo* studies and a moderate level of evidence from *in vitro* studies, thus these findings not supported any upgrading or downgrading modification of the preliminary hazard classification.

Thus the final hazard identification conclusion was that p,p'-DDT and p,p'-DDE are “presumed” to be obesogenic in humans, based on a moderate level of human evidence, moderate level of *in vivo* evidence, a moderate level of evidence from secondary *in vivo* outcomes and *in vitro* studies that supported the biological plausibility of the association.

5. DATA FORMS

Human studies

Author
Year
Funding source
Conflict of interests
Study population name/description
Dates of study and sampling time frame
Geography (country, region,)
Demographics
Gender
Age
Race
N (Number of participants
Recruitment strategy
Inclusion/exclusion criteria
Study design
Study design -Extended
Length of follow-up
Health outcome
Diagnostic or methods used to measure health outcome
Confounders or modifying factors
Substance name and CAS number
Exposure assessment
Units
Methodological details for exposure assessment
Statistical methods
Exposure levels
Statistical findings
Statistical power
Observations on dose response

Animal studies

Author
Title
Year
Funding source
Conflict of interests
Sex
Species
Strain
Source of animals
Age or life-stage at start of dosing and at health outcome assessment
Diet and husbandry information
Chemical name and CAS number
Source of chemical
Purity of chemical
Vehicle used for exposed animals
Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
Other dose-related details such as administered dose levels was verified by measurement, information on internal dosimetry
Route of administration
Duration and frequency of dosing
Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
Guideline compliance
Number of animals per group
Randomization procedure, allocation concealment, blinding during outcome assessment
Method to control for litter effects in developmental studies
Use of negative controls and whether controls were untreated, vehicle-treated, or both
Report on data from positive controls; was expected response observed?
Endpoint
Diagnostic or method to measure endpoint
Statistical methods
Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results.

***In vitro* studies**

Author
Year
Cell line, cell type, or tissue
Chemical name and CAS number
Concentration levels (as presented and converted to μM when possible)
Vehicle used for experimental/control conditions
Endpoint or assay target
Name and source of assay kit
Diagnostic or method to measure endpoint (e.g., reporter gene)
Statistical methods
No Observed Effect Concentration (NOEC), Lowest Observed Effect Concentration (LOEC), statistical significance of other concentration levels, AC50, or other estimates of effect presented in paper. Note: The NOEC and LOEC are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEC does not necessarily mean zero response.
Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)

6. INSTRUCTIONS TO ASSESS THE RISK OF BIAS OF HUMAN EPIDEMIOLOGICAL STUDIES

CONFOUNDING BIAS

1. Did the study design or analysis account for important confounding and modifying variables?

PERFORMANCE BIAS

2. Did researchers adjust or control for other exposures that are anticipated to bias results?

ATTRITION/EXCLUSION BIAS

3. Were outcome data incomplete due to attrition or exclusion from analysis?

DETECTION BIAS

4. Were the outcome assessors blinded to study group or exposure level?
5. Can we be confident in the exposure characterization?
6. Can we be confident in the outcome assessment?

SELECTIVE REPORTING BIAS

7. Were all measured outcomes reported?

CONFLICT OF INTEREST

8. Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?

General Answer Format:

(--) **Definitely Low risk of bias:** There is direct evidence of low risk of bias practices in the form of an explicit statement from the study report or through contacting the authors

(-) **Probably Low risk of bias:** Low risk of bias practice can be inferred from study report (“indirect evidence”) **OR** it is deemed by the risk of bias evaluator that deviations from definitely low risk of bias practices would not appreciably bias results, including consideration of direction and magnitude of bias.

(+) **Probably High risk of bias:** There is indirect evidence of high risk of bias practices **OR** there is insufficient information provided about relevant risk of bias practices to infer.

(++) **Definitely High risk of bias:** There is direct evidence of high risk of bias practices

Confounding bias 1. Did the study design or analysis account for important confounding and modifying variables?

Based on a preliminary search on the literature we elaborated a directed acyclic graph (DAG) to identify relevant covariates (Figure S7). A large list of individual and maternal variables was identified to be associated with exposure to DDTs and/or obesity. We defined as key covariables, those relevant variables identified in the DAG but also explaining more than 10 % of overall variance in the modes of published studies. We selected as potential key confounders maternal BMI, maternal smoking and sex.

Definitely Low risk of bias:

There is direct evidence that appropriate adjustments or explicit considerations were made for primary covariates and confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, case matching, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included.

Probably Low risk of bias:

There is indirect evidence that appropriate adjustments were made for most primary covariates and confounders OR it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results.

Probably High risk of bias:

There is indirect evidence that the distribution of primary covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses OR there is insufficient information provided about the distribution of known confounders.

Definitely High risk of bias:

There is direct evidence that the distribution of primary covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses.

Performance bias 2. Did researchers adjust or control for other exposures that are anticipated to bias results?

The direction of the bias (towards or away from the null) will differ based on the nature of unintended exposure. For example, in a human study if the exposed group lives at a Superfund site they may be exposed to high levels of other environmental contaminants that, if not accounted for, may bias results away from the null (towards larger effects sizes).

It is understood in environmental health that people are exposed to complex mixtures of environmental contaminants and other types of exposures that make it difficult to establish chemical-specific associations. Thus, we will not penalize studies if other exposures are not adjusted or controlled for in most cases. For some projects exceptions may include studies where levels of other chemicals aside from the chemical of interest are likely to be high, such as in occupational cohorts or contaminated regions (e.g., Superfund sites). For some health outcomes, consideration of additional therapies, including medications, may also be appropriate.

Definitely Low risk of bias:

There is direct evidence that other exposures anticipated to bias results were not present or were appropriately adjusted for. For occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered.

Probably Low risk of bias:

There is indirect evidence that other co-exposures anticipated to bias results were not present or were appropriately adjusted for OR it is deemed that co-exposures present would not appreciably bias results. Note, as discussed above, this includes insufficient information provided on co-exposures in general population studies.

Probably High risk of bias:

There is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for OR there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated.

Definitely High risk of bias:

Co, CrSe, CaS: There is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

Attrition/exclusion bias 3. Were outcome data incomplete due to attrition or exclusion from analysis?

Incomplete outcome data includes loss due to attrition (nonresponse, dropout, or loss to follow-up) or exclusion from analyses. The degree of bias resulting from incomplete outcome data depends on the reasons that outcomes are missing, the amount and distribution of missing data across groups, and the potential association between outcome values and likelihood of missing data (Higgins and Green 2011). The risk of bias from incomplete outcome data can be reduced if study authors address the problem in their analyses (e.g., intention to treat analysis and imputation).

Differential or overall attrition because of nonresponse, dropping out, loss to follow-up, and exclusion of participants can introduce bias when missing outcome data are related to both exposure/treatment and outcome. Those who drop out of the study or who are lost to follow-up may be systematically different from those who remain in the study. Attrition or exclusion bias can potentially change the collective (group) characteristics of the relevant groups and their observed outcomes in ways that affect study results by confounding and spurious associations (Viswanathan et al. 2012). This risk of bias item is recommended to assess observational human studies (Viswanathan et al. 2012). However, concern over bias from incomplete outcome data is mainly theoretical and most studies that have looked at whether aspects of missing data are associated with magnitude of effect estimates have not found clear evidence of bias (reviewed in Higgins and Green 2011).

Definitely Low risk of bias:

There is direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study. Acceptable handling of subject attrition includes: very little missing outcome data; reasons for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; OR missing data have been imputed using appropriate methods, AND characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.

Probably Low risk of bias:

There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study OR it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from

those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable.

Probably High risk of bias:

There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed OR there is insufficient information provided about numbers of subjects lost to follow-up.

Definitely High risk of bias:

There is direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.

Detection bias 4. Were the outcome assessors blinded to study group or exposure level?

Blinding requires that outcome assessors do not know the study group or exposure level of the human subject or animal when the outcome was assessed.

If outcome assessors are not blinded to the study group or exposure level it could bias the outcome assessment, so this is a recommended risk of bias element for controlled trials and observational studies (Higgins and Green 2011; Viswanathan et al. 2012).

Without distinguishing between the different stages of blinding during the conduct of a study, lack of blinding in randomized trials has been empirically shown to be associated with larger estimations of intervention effects (on average a 9% increase in an odds ratio) (Pildal et al. 2007). Schulz et al. (1995) analyzed 250 controlled trials and found that studies that were not double-blinded had a 17% larger estimation of treatment effect, on average. In trials with more subjective outcomes, more bias has been observed with lack of blinding (Wood et al. 2008), indicating that blinding outcome assessors could be more important for these effects.

For some exposures, it is not possible to entirely blind outcome assessors, particularly if subjects are self-reporting outcomes. However, adherence to a strict study protocol can reduce the risk of bias. In practice, successful blinding cannot be ensured, as it can be compromised for most interventions. In some cases the treatment may have side effects possibly allowing the participant to detect which intervention they received, unless the study compares interventions with similar side effects or uses an active placebo (Boutron et al. 2006).

Definitely Low risk of bias:

There is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes.

Probably Low risk of bias:

There is indirect evidence that the outcome assessors were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes OR it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome lack of blinding is unlikely to bias a particular outcome).

Probably High risk of bias:

There is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome) OR there is insufficient information provided about blinding of outcome assessors.

Definitely High risk of bias:

There is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

Detection bias 5. Can we be confident in the exposure characterization?

Detection bias can be minimized by using valid and reliable exposure measures applied consistently across groups consistently assessed (i.e., under the same method and time-frame). For example, studies relying on indirect measures of exposure (e.g., self-report) may be rated as having a higher risk of bias than studies that directly measure exposure (e.g., measurement of the chemical in air or measurement of the chemical in blood, plasma, urine, etc.).

GC methodology provides high resolution and a reproducibility of retention time, which is ideal for distinguishing between the p,p'- and o,p'-isomers of the compounds, especially when using GC capillary

columns (Mukherjee and Gopal 1996). Both the GC/ECD and GC/MS analytical methods are suitable for the analysis of DDT, DDE, and DDD. However, the GC/ECD method typically provides greater detection sensitivity, whereas the GC/MS method has the advantage of providing qualitative information to determine the specificity of the analysis. The results may be reported on a lipid or fat basis (i.e., ng DDT/g lipids). By reporting monitoring studies of DDT on a lipid basis, variability in results due to variability in fat content is reduced (McKinney et al. 1984; Phillips et al. 1989).

Definitely Low risk of bias:

Use of robust analytical methodology providing enough information on method performance parameters and quality assessment procedures. There is direct evidence that most data points for the DDT isomers, specially p,p'-DDE are above the level of quantitation (LOQ) for the assay; AND the study utilized spiked samples to confirm assay performance and the stability of DDTs in biological samples was appropriately addressed.

Use of a single measurement in large sample size studies such as NHANES is less of a issue because the number of participants offsets potential concern for differential exposure misclassification. We will not downgrade if a study did not follow these preferred practices.

Probably Low risk of bias:

There is indirect evidence about the use of robust analytical methodology providing enough information on method performance parameters and quality assessment procedures. There is indirect evidence that most data points for the DDT isomers, specially p,p'-DDE are above the level of quantitation (LOQ) for the assay; AND the study utilized spiked samples to confirm assay performance and the stability of DDTs in biological samples was appropriately addressed

Probably High risk of bias:

There is indirect or direct evidence that most individual data points for the DDTS, especially for p,p'-DDE are below the level of quantitation (LOQ) for the assay; OR use of questionnaire items that are not supported by results of biomonitoring studies OR job description for occupational studies that are not supported by information on levels in the work environment or results of biomonitoring studies

Definitely High risk of bias:

The authors did not report the methods used to assess exposure and this information could not be obtained through author query; OR there is evidence of self-report of exposure.

Detection bias 6. Can we be confident in the outcome assessment?

Blinding of outcome assessors is a widely recommended risk-of-bias element for controlled trials and observational studies (Higgins and Green 2011, Viswanathan et al. 2012, Sterne et al. 2014). For human studies blinding of the subject to exposure levels should also be considered. For example, a subject's knowledge of their own exposure levels would represent an increased risk of bias for self-reported outcomes relative to clinically measured outcomes.

Definitively Low risk of bias:

There is direct evidence that anthropometric measurements were performed by clinicians or well trained personnel using gold-standard methods. Classification of subjects was established by official charts

Probably Low risk of bias:

There is indirect evidence that anthropometric measurements were performed by clinicians or well trained personnel using acceptable methods.

Probably High risk of bias:

There is indirect evidence that the outcome assessment method is an insensitive methodology, the authors did not validate the methods used, or the length of follow up differed by study group OR there is insufficient information provided about validation of outcome assessment method.

Definitely High risk of bias:

There is direct evidence that the outcome assessment method is an insensitive methodology (e.g. Self-reported questionnaires), or the length of follow up differed by study group.

Selective reporting bias 7. Were all measured outcomes reported?

Selective reporting of results is a recommended element of assessing risk of bias (Guyatt et al. 2011; Higgins et al. 2011; IOM 2011; Viswanathan et al. 2012). Selective reporting is present if pre-specified outcomes are not reported or incompletely reported. It is likely widespread and difficult to assess with confidence for most studies unless the study protocol is available. Selective reporting bias can be assessed by comparing the “methods” and “results” section of the paper, and by considering outcomes measured in the context of knowledge in the field. Abstracts of presentations relating to the study may contain information about outcomes not subsequently mentioned in publications. Selective reporting bias should be suspected if the study does not report outcomes in the results section that would have been expected

based on the methods, or if a composite score is present without the individual component outcomes (Guyatt et al. 2011). It may be useful to pay attention to author affiliations and funding source which can contribute to selective outcome reporting when results are not consistent with expectations or value to the research objectives.

Definitely Low risk of bias:

There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction.

Probably Low risk of bias:

There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported OR analyses that had not been planned at the outset of the study (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and it is deemed that the omitted analyses were not appropriate and selective reporting would not appreciably bias results. This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).

Probably High risk of bias:

There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported OR there is insufficient information provided about selective outcome reporting.

Definitely High risk of bias:

There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect).

Conflict of interest 8. Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?

Definitely Low risk of bias:

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. A conflict of interest statement is provided to indicate the authors have no financial interest and there is evidence of the entities not having a financial interest.

Probably Low risk of bias:

There is insufficient information to permit a judgment of 'YES', for example there is no conflict of interest statement denying financial interests, but there is evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'YES'.

Probably high risk of bias:

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'NO'.

Definitely High risk of bias:

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study.

7. INSTRUCTIONS TO ASSESS THE RISK OF BIAS OF *IN VIVO* STUDIES

SEQUENCE GENERATION

1. Was the allocation sequence adequately generated?

ALLOCATION CONCEALMENT

2. Was allocation adequately concealed?

BLINDING OF PERSONNEL AND OUTCOME ASSESSORS

3. Was knowledge of allocated interventions adequately prevented?

INCOMPLETE OUTCOME DATA

4. Were incomplete outcome data adequately addressed?

SELECTIVE OUTCOME REPORTING

5. Were study reports free of selective outcome reporting?

OTHER POTENTIAL THREATS TO VALIDITY

6. Was study free of other problems regarding risk of bias?

CONFLICT OF INTEREST

7. Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?

SEQUENCE GENERATION

1. Was the allocation sequence adequately generated?

Was the allocation sequence adequately generated?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The investigators describe a random component in the sequence generation process such as:

- Referring to a random number table;
- Using a computer random number generator;
- Coin tossing;
- Shuffling cards or envelopes;
- Throwing dice;
- Drawing of lots.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'YES', but there is indirect evidence that suggests the sequence generation process was random, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'NO', but there is indirect evidence that suggests a non-random component in the sequence generation process, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The investigators describe a non-random component in the sequence generation process or that a random component was not used. Usually, the description would involve some systematic, non-random approach, for example:

- Sequence generated by date of birth;
- Sequence generated by some rule based on date (or day) of arrival at facility;
- Sequence generated by some rule based on record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of animals, for example:

- Allocation by judgment of the investigator;
- Allocation by judgment of the intervention.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

ALLOCATION CONCEALMENT

2. Was allocation adequately concealed?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

Investigators could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:

- Sequentially numbered treatment containers of identical appearance to control; or
- Sequentially numbered prepared route of administration (e.g., pre-prepared water dosed with chemical) of identical appearance; or
- Sequentially numbered, opaque, sealed envelopes.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about allocation concealment to permit a judgment of 'YES', but there is indirect evidence that suggests the allocation was adequately concealed, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about allocation concealment to permit a judgment of 'NO', but there is indirect evidence that suggests the allocation was not adequately concealed, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Investigators handling experimental animals could possibly foresee assignments and thus introduce selection bias, such as allocation based on:

- Using an open random allocation schedule (e.g. a list of random numbers); or
- Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); or
- Alternation or rotation; or
- Non-random and known criteria, such as date of birth; or
- Record number; or
- Any other explicitly unconcealed procedure.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

BLINDING OF PERSONNEL AND OUTCOME ASSESSORS

3. Was knowledge of allocated interventions adequately prevented?

Was knowledge of the allocated interventions adequately prevented during the study?

Criteria for a judgment of ‘YES’ (i.e. low risk of bias):

Any one of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; or
- Blinding of key study personnel ensured, and unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias.

Criteria for the judgment of ‘PROBABLY YES’ (i.e. probably low risk of bias):

There is insufficient information about blinding to permit a judgment of ‘YES’, but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of ‘YES’.

Criteria for the judgment of ‘PROBABLY NO’ (i.e. probably high risk of bias):

There is insufficient information about blinding to permit a judgment of ‘NO’, but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of ‘NO’.

Criteria for the judgment of ‘NO’ (i.e. high risk

of bias): Any one of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken; or
- Some key study personnel were not blinded, and the non-blinding of others likely to introduce bias.

Criteria for the judgment of ‘NOT APPLICABLE’ (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

INCOMPLETE OUTCOME DATA

4. Were incomplete outcome data adequately addressed?

Were incomplete outcome data adequately addressed?

Criteria for a judgment of ‘YES’ (i.e. low risk of bias):

The number of animals assessed for outcome of interest is reported and data is provided indicating adequate follow up of all treated animals. Additional information provided by authors should be considered when making risk of bias judgments about incomplete outcome data.

Additionally, any one of the following:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); or
- Missing outcome data is provided and is balanced in numbers across intervention groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a biologically relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a biologically relevant impact on observed effect size; or
- Missing data have been imputed using appropriate statistical methods.

Criteria for the judgment of ‘PROBABLY YES’ (i.e. probably low risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of ‘YES’, but there is indirect evidence that suggests incomplete outcome data were adequately addressed, as described by the criteria for a judgment of ‘YES’.

Criteria for the judgment of ‘PROBABLY NO’ (i.e. probably high risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of ‘NO’, but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of ‘NO’.

Criteria for the judgment of ‘NO’ (i.e. high risk of bias):

The number of animals allocated not reported and no data is provided to indicate that there was adequate follow up of all treated animals. Additionally, any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- ‘As-treated’ analysis done with substantial departure of the intervention received from that assigned at randomization; or

- Potentially inappropriate application of simple imputation.

Criteria for the judgment of ‘NOT APPLICABLE’ (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

SELECTIVE OUTCOME REPORTING

5. Were study reports free of selective outcome reporting?

Criteria for a judgment of ‘YES’ (i.e. low risk of bias):

All of the study’s pre-specified (primary and secondary) outcomes outlined in the methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, including the number of animals analyzed for outcomes of interest. Additional information provided by authors should be considered when making risk of bias judgments for selective outcome reporting.

Criteria for the judgment of ‘PROBABLY YES’ (i.e. probably low risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of ‘YES’, but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of ‘YES’.

Criteria for the judgment of ‘PROBABLY NO’ (i.e. probably high risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of ‘NO’, but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of ‘NO’.

Criteria for the judgment of ‘NO’ (i.e. high risk of bias):

The study was not free of selective reporting. The following should be considered:

- Authors did not report numbers analyzed for outcomes of interest; or
- Not all of the study’s pre-specified primary outcomes (as outlined in the protocol, title, abstract, and/or introduction) that are of interest in the review have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); or
- One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; or

- The study report fails to include results for a key outcome that would be expected to have been reported for such a study.

Criteria for the judgment of ‘NOT APPLICABLE’ (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

OTHER POTENTIAL THREATS TO VALIDITY

6. Was study free of other problems regarding risk of bias?

We considered in this section other sources of potential bias not considered in the previous sections. Example: Failure to statistically or experimentally adjust for litter in an animal study with a developmental outcome. The direction of the bias is away from the null towards a larger effect size.

Criteria for a judgment of ‘YES’ (i.e. low risk of bias):

The study appears to be free of other sources of bias.

Criteria for the judgment of ‘PROBABLY YES’ (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of ‘YES’, but there is indirect evidence that suggests the study was free of other threats to validity, as described by the criteria for a judgment of ‘YES’.

Criteria for the judgment of ‘PROBABLY NO’ (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of ‘NO’, but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of ‘NO’.

Criteria for the judgment of ‘NO’ (i.e. high risk of bias):

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used;
- Stopped early due to some data-dependent process (including a formal-stopping rule);
- Had extreme baseline imbalance (improper control group);
- Has been claimed to have been fraudulent;
- The conduct of the study is affected by interim results (e.g. recruiting additional animals from a subgroup showing more benefit);
- There is deviation from the study protocol in a way that does not reflect typical practice (e.g. post hoc stepping-up of doses to exaggerated levels);

- There is pre-randomization administration of an intervention that could enhance or diminish the effect of a subsequent, randomized, intervention; inappropriate administration of an intervention (or co-intervention);
- Occurrence of ‘null bias’ due to interventions being insufficiently well delivered or overly wide inclusion criteria for animals;
- An insensitive instrument is used to measure outcomes (which can lead to under-estimation of both beneficial and harmful effects);
- Selective reporting of subgroups;
- Had some other problem.

CONFLICT OF INTEREST

7. Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?

Criteria for a judgment of ‘YES’ (i.e. low risk of bias):

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. A conflict of interest statement is provided to indicate the authors have no financial interest and there is evidence of the entities not having a financial interest. Examples of this evidence include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;
- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of ‘PROBABLY YES’ (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of 'YES', for example there is no conflict of interest statement denying financial interests, but there is evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;

8. REFERENCES

- Agay-Shay K, Martinez D, Valvi D, Garcia-Esteban R, Basagana X, Robinson O, et al. 2015. Exposure to endocrine-disrupting chemicals during pregnancy and weight at 7 years of age: A multi-pollutant approach. *Environ Health Perspect* 123:1030-1037.
- Ahdaya SM, Shah PV, Guthrie FE. 1976. Thermoregulation in mice treated with parathion, carbaryl, or ddt. *Toxicol Appl Pharmacol* 35:575-580.
- Allison DB, Downey M, Atkinson RL, Billington CJ, Bray GA, Eckel RH, et al. 2008. Obesity as a disease: A white paper on evidence and arguments commissioned by the council of the obesity society. *Obesity (Silver Spring, Md)* 16:1161-1177.
- Appleton BS, Shriver CN, Arnrich L, Hathcock JN. 1981. Effects of 3-methylcholanthrene and ddt on cholesterol 7 alpha-hydroxylase in rats. *Drug-nutrient interactions* 1:15-21.
- Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M, et al. 2015. Global burden of cancer attributable to high body-mass index in 2012: A population-based study. *The Lancet Oncology* 16:36-46.
- ATSDR. 2002. Toxicological profile for ddt, dde, and ddd. Atlanta, .
- Beard J. 2006. Ddt and human health. *Sci Total Environ* 355:78-89.
- Benton MJ, Nimrod AC, Benson WH. 1994. Evaluation of growth and energy storage as biological markers of ddt exposure in sailfin mollies. *Ecotoxicol Environ Saf* 29:1-12.
- Billionnet C, Sherrill D, Annesi-Maesano I. 2012. Estimating the health effects of exposure to multi-pollutant mixture. *Annals of epidemiology* 22:126-141.
- Cupul-Uicab LA, Hernandez-Avila M, Terrazas-Medina EA, Pennell ML, Longnecker MP. 2010. Prenatal exposure to the major ddt metabolite 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (dde) and growth in boys from Mexico. *Environmental research* 110:595-603.
- Cupul-Uicab LA, Klebanoff MA, Brock JW, Longnecker MP. 2013. Prenatal exposure to persistent organochlorines and childhood obesity in the U.S. Collaborative perinatal project. *Environmental Health Perspectives* 121:1103-1109.
- Chapados NA, Casimiro C, Batal M, Blais JM, Haman F, Robidoux MA, et al. 2011. Increased proliferative effect of organochlorine compounds on human preadipocytes. *Canadian Journal of Diabetes* 35:213.
- Darsie J, Gosha SK, Holman RT. 1976. Induction of abnormal fatty acid metabolism and essential fatty acid deficiency in rats by dietary ddt. *Archives of Biochemistry and Biophysics* 175:262-269.
- Delvaux I, Van Cauwenberghe J, Den Hond E, Schoeters G, Govarts E, Nelen V, et al. 2014. Prenatal exposure to environmental contaminants and body composition at age 7-9 years. *Environmental research* 132:24-32.
- Faroon O, Harris MO. 2002. Toxicological profile for ddt, dde, and ddd. United States:Syracuse Research Corporation.
- Gaskins AJ, Schisterman EF. 2009. The effect of lipid adjustment on the analysis of environmental contaminants and the outcome of human health risks. *Methods in molecular biology (Clifton, NJ)* 580:371-381.
- Gladden BC, Klebanoff MA, Hediger ML, Katz SH, Barr DB, Davis MD, et al. 2004. Prenatal ddt exposure in relation to anthropometric and pubertal measures in adolescent males. *Environmental Health Perspectives* 112:1761-1767.
- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. 2011a. Grade guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol* 64:1283-1293.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. 2011b. Grade guidelines: 8. Rating the quality of evidence--indirectness. *Journal of clinical epidemiology* 64:1303-1310.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. 2011c. Grade guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of clinical epidemiology* 64:1294-1302.

Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. 2011d. Grade guidelines: 5. Rating the quality of evidence--publication bias. *Journal of clinical epidemiology* 64:1277-1282.

Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. 2011e. Grade guidelines: 9. Rating up the quality of evidence. *Journal of clinical epidemiology* 64:1311-1316.

Higgins JPT, Green S. 2011. *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Available: www.cochrane-handbook.org [accessed 15-08-2015 2015].

Howell G, Mangum L. 2011. Exposure to bioaccumulative organochlorine compounds alters adipogenesis, fatty acid uptake, and adipokine production in nih3t3-l1 cells. *Toxicology in vitro : an international journal published in association with BIBRA* 25:394-402.

Howell GE, Meek E, Kilic J, Mohns M, Mulligan C, Chambers JE. 2014. Exposure to p,p'-dichlorodiphenyldichloroethylene (dde) induces fasting hyperglycemia without insulin resistance in male c57bl/6h mice. *Toxicology* 320:6-14.

Howell GE, Mulligan C, Meek E, Chambers JE. 2015. Effect of chronic p,p'-dichlorodiphenyldichloroethylene (dde) exposure on high fat diet-induced alterations in glucose and lipid metabolism in male c57bl/6h mice. *Toxicology* 328:112-122.

Hoyer BB, Ramlau-Hansen CH, Henriksen TB, Pedersen HS, Goralczyk K, Zvezdai V, et al. 2014. Body mass index in young school-age children in relation to organochlorine compounds in early life: A prospective study. *International Journal of Obesity* 38:919-925.

Ibrahim MM, Fjaere E, Lock EJ, Naville D, Amlund H, Meugnier E, et al. 2011. Chronic consumption of farmed salmon containing persistent organic pollutants causes insulin resistance and obesity in mice. *PLoS one* 6:e25170.

Ishikawa T, Graham JL, Stanhope KL, Havel PJ, La Merrill MA. 2015. Effect of ddt exposure on lipids and energy balance in obese sprague-dawley rats before and after weight loss. *Toxicology Reports* 2:990-995.

JECFA JFWECofA. 2000. *Guidelines for the preparation of toxicological working papers for the joint fao/who expert committee on food additives*. Geneva, Switzerland.

Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Systematic review of human evidence for pfoa effects on fetal growth. *Environ Health Perspect* 122:1028-1039.

Jokinen E. 2015. Obesity and cardiovascular disease. *Minerva pediatrica* 67:25-32.

Karmaus W, Osuch JR, Eneli I, Mudd LM, Zhang J, Mikucki D, et al. 2009. Maternal levels of dichlorodiphenyl-dichloroethylene (dde) may increase weight and body mass index in adult female offspring. *Occup Environ Med* 66:143-149.

Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for pfoa effects on fetal growth. *Environ Health Perspect* 122:1015-1027.

Kulig M, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A, et al. 2012. [grade guidelines: 6. Rating the quality of evidence: Imprecision]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 106:677-688.

La Merrill M, Emond C, Kim MJ, Antignac JP, Le Bizec B, Clement K, et al. 2013. Toxicological function of adipose tissue: Focus on persistent organic pollutants. *Environmental Health Perspectives* 121:162-169.

La Merrill M, Karey E, Moshier E, Lindtner C, La Frano MR, Newman JW, et al. 2014. Perinatal exposure of mice to the pesticide ddt impairs energy expenditure and metabolism in adult female offspring. *PLoS one* 9:e103337.

Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Integration of animal and human evidence for pfoa effects on fetal growth. *Environ Health Perspect* 122:1040-1051.

Lee DH, Lind L, Jacobs DR, Salihovic S, van Bavel B, Lind PM. 2012. Associations of persistent organic pollutants with abdominal obesity in the elderly: The prospective investigation of the vasculature in uppsala seniors (pivus) study. *Environment international* 40:170-178.

Li D, Longnecker MP, Dunson DB. 2013. Lipid adjustment for chemical exposures: Accounting for concomitant variables. *Epidemiology* 24:921-928.

Loomis D, Guyton K, Grosse Y, El Ghissasi F, Bouvard V, Benbrahim-Tallaa L, et al. 2015. Carcinogenicity of lindane, ddt, and 2,4-dichlorophenoxyacetic acid. *The Lancet Oncology* 16:891-892.

Lopez-Carrillo L, Torres-Sanchez L, Lopez-Cervantes M, Blair A, Cebrian ME, Uribe M. 1999. The adipose tissue to serum dichlorodiphenyldichloroethane (dde) ratio: Some methodological considerations. *Environ Res* 81:142-145.

Mangum LH, Howell GE, Chambers JE. 2015. Exposure to p,p'-dde enhances differentiation of 3t3-l1 preadipocytes in a model of sub-optimal differentiation. *Toxicology letters* 238:65-71.

Moreno-Aliaga MJ, Matsumura F. 2002. Effects of 1,1,1-trichloro-2,2-bis(p-chlorophenyl)-ethane (p,p'-ddt) on 3t3-l1 and 3t3-f442a adipocyte differentiation. *Biochemical Pharmacology* 63:997-1007.

Nagaoka S, Masaki H, Aoyama Y, Yoshida A. 1986. Effects of excess dietary tyrosine or certain xenobiotics on the cholesterol synthesis in rats. *J Nutr* 116:726-732.

Narayan S, Dani HM, Misra UK. 1990. Changes in lipid profiles of liver microsomes of rats following intratracheal administration of ddt or endosulfan. *Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes* 25:243-257.

O'Brien KM, Upson K, Cook NR, Weinberg CR. 2015. Environmental chemicals in urine and blood: Improving methods for creatinine and lipid adjustment. *Environmental health perspectives*.

Ogden CL, Carroll MD, Kit BK, Flegal KM. 2014. Prevalence of childhood and adult obesity in the united states, 2011-2012. *Jama* 311:806-814.

OHAT. 2015a. Ohat risk of bias rating tool for human and animal studies.

OHAT. 2015b. Handbook for conducting a literature-based health assessment using ohat approach for systematic review and evidence integration.

Okazaki Y, Katayama T. 2003. Effects of dietary carbohydrate and myo-inositol on metabolic changes in rats fed 1,1,1-trichloro-2,2-bis (p-chlorophenyl) ethane (ddt). *J Nutr Biochem* 14:81-89.

Okazaki Y, Katayama T. 2008. Dietary inositol hexakisphosphate, but not myo-inositol, clearly improves hypercholesterolemia in rats fed casein-type amino acid mixtures and 1,1,1-trichloro-2,2-bis (p-chlorophenyl) ethane. *Nutrition Research* 28:714-721.

Patel CJ, Cullen MR, Ioannidis JP, Butte AJ. 2012. Systematic evaluation of environmental factors: Persistent pollutants and nutrients correlated with serum lipid levels. *International journal of epidemiology* 41:828-843.

Phillips DL, Pirkle JL, Burse VW, Bernert Jr JT, Henderson LO, Needham LL. 1989. Chlorinated hydrocarbon levels in human serum: Effects of fasting and feeding. *ARCH ENVIRON CONTAM TOXICOL* 18:495-500.

Platt DS, Cockrill BL. 1967. Liver enlargement and hepatotoxicity: An investigation into the effects of several agents on rat liver enzyme activities. *Biochemical Pharmacology* 16:2257-2270.

Platt DS, Cockrill BL. 1969. Biochemical changes in rat liver in response to treatment with drugs and other agents. Ii. Effects of halothane, ddt, other chlorinated hydrocarbons, thioacetamide, dimethylnitrosamine and ethionine. *Biochem Pharmacol* 18:445-457.

Porta M, Jarrod M, Lopez T, Pumarega J, Puigdomenech E, Marco E, et al. 2009. Correcting serum concentrations of organochlorine compounds by lipids: Alternatives to the organochlorine/total lipids ratio. *Environ Int* 35:1080-1085.

Rao TN, Sanyal S, Agarwal N, Subrahmanyam D. 1981. Effect of a single oral dose of ddt on lipid metabolism in rat. *Toxicology Letters* 7:263-266.

Rodriguez-Alcala LM, Sa C, Pimentel LL, Pestana D, Teixeira D, Faria A, et al. 2015. Endocrine disruptor dde associated with a high-fat diet enhances the impairment of liver fatty acid composition in rats. *Journal of agricultural and food chemistry* 63:9341-9348.

Rogan WJ, Chen A. 2005. Health risks and benefits of bis(4-chlorophenyl)-1,1,1-trichloroethane (ddt). *Lancet (London, England)* 366:763-773.

Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, et al. 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environment international*.

Schisterman EF, Whitcomb BW, Louis GM, Louis TA. 2005. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environ Health Perspect* 113:853-857.

Schisterman EF, Cole SR, Platt RW. 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 20:488-495.

Skinner MK, Manikkam M, Tracey R, Guerrero-Bosagna C, Haque M, Nilsson EE. 2013. Ancestral dichlorodiphenyltrichloroethane (ddt) exposure promotes epigenetic transgenerational inheritance of obesity. *BMC Med* 11.

Smith D. 1999. Worldwide trends in ddt levels in human breast milk. *International journal of epidemiology* 28:179-188.

Speakman JR, O'Rahilly S. 2012. Fat: An evolving issue. *Disease models & mechanisms* 5:569-573.

Strong AL, Shi Z, Strong MJ, Miller DF, Rusch DB, Buechlein AM, et al. 2015. Effects of the endocrine-disrupting chemical ddt on self-renewal and differentiation of human mesenchymal stem cells. *Environmental Health Perspectives* 123:42-48.

Takayama S, Sieber SM, Dalgard DW, Thorgeirsson UP, Adamson RH. 1999. Effects of long-term oral administration of ddt on nonhuman primates. *Journal of cancer research and clinical oncology* 125:219-225.

Tang-Peronard JL, Heitmann BL, Andersen HR, Steuerwald U, Grandjean P, Weihe P, et al. 2014. Association between prenatal polychlorinated biphenyl exposure and obesity development at ages 5 and 7 y: A prospective cohort study of 656 children from the faroe islands. *American Journal of Clinical Nutrition* 99:5-13.

Tang-Peronard JL, Jensen TK, Andersen HR, Ried-Larsen M, Grontved A, Andersen LB, et al. 2015. Associations between exposure to persistent organic pollutants in childhood and overweight up to 12 years later in a low exposed danish population. *Obesity facts* 8:282-292.

Taxvig C, Boberg J, Nellemann C, Marie A. 2012. Effect of environmental pollutants on obesity parameters in vitro. *Toxicology Letters* 211:S49.

UNEP. 2010. Report of the expert group on the assessment of the production and use of ddt and its alternatives for disease vector control. Third meeting. Geneva:United Nations Environment Programme.

Vafeiadi M, Georgiou V, Chalkiadaki G, Rantakokko P, Kiviranta H, Karachaliou M, et al. 2015. Association of prenatal exposure to persistent organic pollutants with obesity and cardiometabolic traits in early childhood: The reha mother-child cohort (crete, greece). *Environ Health Perspect*.

Valvi D, Mendez MA, Martinez D, Grimalt JO, Torrent M, Sunyer J, et al. 2012. Prenatal concentrations of polychlorinated biphenyls, dde, and ddt and overweight in children: A prospective birth cohort study. *Environmental Health Perspectives* 120:451-457.

van den Berg H, Zaim M, Yadav RS, Soares A, Ameneshewa B, Mnzava A, et al. 2012. Global trends in the use of insecticides to control vector-borne diseases. *Environ Health Perspect* 120:577-582.

Warner M, Wesselink A, Harley KG, Bradman A, Kogut K, Eskenazi B. 2014. Prenatal exposure to dichlorodiphenyltrichloroethane and obesity at 9 years of age in the chamacos study cohort. *Am J Epidemiol* 179:1312-1322.

Westlake GE, Bunyan PJ, Stanley PI, Walker CH. 1979. The effects of 1,1-di(p-chlorophenyl)-2-chloroethylene on plasma enzymes and blood constituents in the japanese quail. *Chemico-biological interactions* 25:197-210.

WHO. 2014. Global status report on noncommunicable diseases 2014. Geneva, Switzerland.

Woodruff TJ, Sutton P. 2014. The navigation guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. *Environ Health Perspect* 122:1007-1014.

You L, Gazi E, Archibeque-Engle S, Casanova M, Conolly RB, Heck HA. 1999. Transplacental and lactational transfer of p,p'-dde in sprague-dawley rats. *Toxicol Appl Pharmacol* 157:134-144.