Taylor & Francis
Taylor & Francis Group

Journal of Toxicology and Environmental Health, Part B
Critical Reviews

ISSN: 1093-7404 (Print) 1521-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/uteb20

Systematic reviews and meta-analyses of human and animal evidence of prenatal diethylhexyl phthalate exposure and changes in male anogenital distance

David C. Dorman, Weihsueh Chiu, Barbara F. Hales, Russ Hauser, Kamin J. Johnson, Ellen Mantus, Susan Martel, Karen A. Robinson, Andrew A. Rooney, Ruthann Rudel, Sheela Sathyanarayana, Susan L. Schantz & Katrina M. Waters

To cite this article: David C. Dorman, Weihsueh Chiu, Barbara F. Hales, Russ Hauser, Kamin J. Johnson, Ellen Mantus, Susan Martel, Karen A. Robinson, Andrew A. Rooney, Ruthann Rudel, Sheela Sathyanarayana, Susan L. Schantz & Katrina M. Waters (2018): Systematic reviews and meta-analyses of human and animal evidence of prenatal diethylhexyl phthalate exposure and changes in male anogenital distance, Journal of Toxicology and Environmental Health, Part B, DOI: 10.1080/10937404.2018.1505354

To link to this article: https://doi.org/10.1080/10937404.2018.1505354

View supplementary material

Published online: 10 Sep 2018.

Submit your article to this journal

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=uteb20

JOURNAL OF TOXICOLOGY AND ENVIRONMENTAL HEALTH, PART B
https://doi.org/10.1080/10937404.2018.1505354

Taylor & Francis
Taylor & Francis Group

Check for updates

Systematic reviews and meta-analyses of human and animal evidence of prenatal diethylhexyl phthalate exposure and changes in male anogenital distance

David C. Dorman[a], Weihsueh Chiu[b], Barbara F. Hales[c], Russ Hauser[d], Kamin J. Johnson[e], Ellen Mantus[f], Susan Martel[f], Karen A. Robinson[g], Andrew A. Rooney[h], Ruthann Rudel[i], Sheela Sathyanarayana[j], Susan L. Schantz[k], and Katrina M. Waters[l]

[a]Department of Molecular and Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA; [b]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University, College Station, TX, USA; [c]Department of Pharmacology and Therapeutics, McGill University, Montreal, Quebec, Canada; [d]Department of Environmental Health and Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [e]Predictive Safety Center, The Dow Chemical Company, Midland, MI, USA; [f]Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine, Washington, DC, USA; [g]Department of Medicine, Johns Hopkins University, Baltimore, MD, USA; [h]Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA; [i]Silent Spring Institute, Newton, MA, USA; [j]Department of Pediatrics, University of Washington, Seattle Children's Research Institute, Seattle, WA, USA; [k]Department of Comparative Biosciences, College of Veterinary Medicine and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [l]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

ABSTRACT

Male reproductive alterations found in animals and humans following *in utero* phthalate exposure include decreased anogenital distance (AGD) and other reproductive-tract malformations. The aim of this investigation was to conduct systematic reviews of human and animal evidence of the effect of *in utero* exposure to diethylhexyl phthalate (DEHP) on anogenital distance (AGD) in males. PubMed, Embase, and Toxline were searched for relevant human and experimental animal studies on August 15, 2016. Search results were screened for relevance, and studies that met the inclusion criteria were evaluated for quality and data extracted for analysis. Confidence in the human and animal bodies of evidence was assessed and hazard conclusions reached by integrating evidence streams. The search yielded 6 relevant human studies and 19 animal studies. Meta-analysis of 5 human observational prospective cohort studies showed that increased maternal urinary concentrations of DEHP metabolites were associated with decreased AGD in boys (−4.07 [CI, −6.49 to −1.66] % decrease per $\log_{10}$ rise in DEHP metabolites). Meta-analysis and meta-regression of the 19 experimental animal studies found reduced AGD with DEHP treatment, with a dose-response gradient, and with heterogeneity explained by species and strain. There is a moderate level of evidence from human investigations and a high level of data from animal studies that *in utero* exposure to DEHP decreases AGD. Based upon the available human and animal evidence, and consideration of mechanistic data, DEHP is *presumed* to be a reproductive hazard to humans on the basis of effects on AGD.

KEYWORDS

Reproductive toxicity; phthalate; anogenital distance; epidemiology

Introduction

Phthalates are widely used in a variety of consumer products and human exposure to phthalates occurs following ingestion, dermal exposure, or inhalation (CDC 2009; Lioy et al. 2015). Biomonitoring efforts performed by National Health and Nutrition Examination Survey (NHANES) and others generally rely on the measurement of urinary phthalate metabolite concentrations (Howdeshell et al. 2017; Johns et al. 2015). Phthalates cross the placenta (Fennell et al. 2004), and phthalates and their metabolites have been measured in amniotic fluid (Silva et al. 2004; Calafat et al. 2006; Huang et al. 2009). Transplacental phthalate delivery may lead to adverse developmental effects in animals and may exert similar effects in humans (Gray et al. 2000). In the rat, alterations in male reproductive-tract

CONTACT David C. Dorman david_dorman@ncsu.edu Department of Molecular Biomedical Sciences, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uteb.

Supplemental data for this article can be access on the publisher's website.

© 2018 Taylor & Francis

development are one of the most sensitive health outcomes of *in utero* phthalate exposure (CHAP 2014; NRC 2008). In rats, most phthalates with ester side chains containing 4–6 carbon atoms are anti-androgenic, whereas some other phthalates (e.g., dimethyl and diethyl phthalate) are not anti-androgenic and were not found to adversely affect male reproductive tract development or function in rats (Furr et al. 2014; Gray et al. 2000).

Male reproductive alterations found in animals following *in utero* phthalate exposure have been referred to as "phthalate syndrome" and include decreased anogenital distance (AGD), decreased sperm count, cryptorchidism, hypospadias, infertility, and other reproductive tract malformations (Gray et al. 2000; NRC 2008). Androgen-dependent development of the male reproductive tract and the androgen-dependence of AGD appear to be well conserved across mammals including humans (Hsieh et al. 2012; Jain and Singal 2013; Thankamony et al. 2014). A hypothesized syndrome in individuals ("testicular dysgenesis syndrome") shares some of the same endpoints as rat phthalate syndrome (NRC 2008; Skakkebaek 2002; Wohlfahrt-Veje, Main, and Skakkebaek 2009); however, the etiology of the proposed human syndrome remains unknown. The United States and the European Union have regulated the use of certain phthalates in children's toys and childcare articles due to concerns about phthalate toxicity (Negev et al. 2018).

The review question and specific aims in the present systematic reviews (SRs) were developed and refined through a series of scoping and problem formulation steps. The present SR focused on in utero exposure to diethylhexyl phthalate (DEHP) and effects on AGD in male offspring because of the weight of the data available on DEHP versus other phthalates, as well as the availability of rat and human data. The present SR was designed to address whether in utero exposure to DEHP affected AGD in male offspring.

Methods

Prior SR activities with DEHP and AGD

This study is an offshoot of a National Academies of Sciences, Engineering, and Medicine (NASEM) report that applied SR methods in the evaluation of low dose endocrine effects (NASEM, 2017). The NASEM report included SRs of the human and animal evidence on the effects of *in utero* exposure to phthalates on male reproductive toxicity based on the following outcome measures: AGD; fetal testosterone levels; and incidence of hypospadias (NASEM 2017). The present study focuses on the portion of the NASEM report that evaluated the effect of DEHP on AGD.

Problem formulation and protocol development

Problem formulation approaches and protocols for the conduct of the SR were developed and peer reviewed in accordance with NASEM review practices. The protocols were based on the method developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) for conducting SR (hereto referred to as the OHAT method) (NTP 2015). The protocols specified the research questions; the literature search strategy; the inclusion/exclusion criteria used for identifying relevant studies; framework for judging the quality of included studies; and plan for data analysis, synthesis and presentation of findings (Stephens et al. 2011). The protocols and other details of methods are available from the full report (NASEM 2017).

Research questions

The following research questions were developed for the human and animal systematic reviews:

- Human: What is the effect of *in utero* exposure to DEHP on AGD in male children?
- Animals: What is the effect of *in utero* exposure to DEHP on AGD in nonhuman male animals?

Population, exposure, comparator, and outcome (PECO) statements

The following PECO statements were developed for the animal and human systematic reviews:

Human:

- Population: Male humans
- Exposure: *In utero* exposure to DEHP. No restrictions based on route of exposure.

Measurements must be based on biomonitoring data (e.g., mono-2-ethylhexyl phthalate [MEHP], mono-[2-ethyl-5-hydroxyhexyl] phthalate [MEHHP], mono-[2-ethyl-5-oxohexyl] phthalate [MEOHP], mono-[2-ethyl-5-caroxypentyl] phthalate [MECPP], sum of DEHP metabolites).

- Comparator: Male humans exposed *in utero* to lower concentrations of DEHP.
- Outcomes: AGD: the measured distance between the anus and the genitals. Typically measured from the anus to the base of the scrotum or the base of the phallus. Other measures that might be used include: (a) anogenital index (AGI): AGD measurement divided by body weight or by the cube root of body weight; (b) anoscrotal distance (ASD): the measured distance between the anus and base of the scrotum; or (c) anopenile distance (APD): the measured distance from the anus to the base of the penis.

Animals:

- Population: Nonhuman male mammals
- Exposure: *In utero* exposure to DEHP. Oral route of exposure.
- Comparator: Male nonhuman mammals exposed *in utero* to different doses of DEHP or vehicle-only treatment.
- Outcomes: See human outcomes above.

Literature searches

A search string employing medical subject heading (MeSH) terms and keyword synonyms was developed. The PubMed search strategy was considered the primary search strategy and provided the basis for the other electronic search strategies. To assist in compiling these MeSH terms, 8 human and 25 animal articles were selected for review to help identify spelling variants. The search strategies addressed each of the following concepts: phthalates (e.g., CAS registry numbers were included in the list of search terms), exposure, species, and outcomes. The search for animal literature was completed using a published search filter to eliminate non-mammalian animals (Hooijmans et al. 2010). The search for literature using humans employed a search filter to identify human studies (Higgins and Green 2011) that was modified to comply with PubMed formatting. Each of the above search concepts were searched together using the Boolean operator "AND." PubMed, Embase, and Toxline databases were searched for investigations on the effects of phthalates on male reproductive tract development on August 15, 2016. Reference lists of eligible studies were also searched.

Screening process

References were independently screened for inclusion criteria at the title and abstract level and at the full-text level by two individuals using a web-based project management tool for tracking studies through the screening process (DistillerSR, Evidence Partners Inc., Ottawa, Canada). References were excluded if they met at least one of the following criteria:

- No original data (e.g., review article, commentary, editorial)
- Study does not include relevant population of interest (e.g., male humans or nonhuman mammals)
- Study does not report phthalate exposure
- No relevant outcomes
- Incomplete information (e.g., conference abstract, meeting poster)
- Not in English and unable to determine eligibility
- Other (explanation required)

Data extraction

Data from the included studies were entered into a web-based visual display software system (HAWC; https://hawcproject.org). One person entered data and a second individual verified the entries. All data entered into HAWC are available at the following links: https://hawcproject.org/assessment/351/ (for the animal assessment) and https://hawcproject.org/assessment/350/(for the human assessment).

Quality assessment of individual studies

Risk of bias (RoB) and other study quality elements were assessed using the OHAT RoB tool (NTP

2015) tailored to the SRs, and involved answering up to nine questions, based on the type of study (Supplemental Table 1). Three elements were considered more important for assessing the quality and potential bias in each data set. The 3 key elements in the human epidemiologic studies included control for confounding, exposure characterization, and outcome assessment (including blinding of outcome assessors). The following variables were considered as key potential confounders and/or effect measure modifiers that were considered in the analyses of the relationship between phthalate exposure and AGD: a measure of weight or body size at exam, a measure of weight or body size at birth, age at exam, and measure of urinary dilution (specific gravity, creatinine, or osmolality) or indication that exposure measure was adjusted for urinary dilution. The quality assessment also included consideration of the exposure characterization in the epidemiologic studies, including the reliability of the analytical chemistry methods used, whether exposure biomarkers were measured in a relevant time-window for the outcome, and whether the analysis accounted for urinary dilution. In animal studies, the key elements included whether or not animals were randomly assigned to treatment groups, outcome assessment (including blinding of outcome assessors to treatment groups), and investigator control for litter effects in the experimental design or statistical approaches (e.g., using the "litter" as the statistical unit of analysis, or utilizing a nested model or another statistical method that accounts for intra-litter correlation). The quality assessment of the animal studies also considered when outcome assessments were performed (i.e., age), characterization of the test chemical, exposure methods, concealment of allocation to study groups, and information regarding attrition and data exclusion. Two-person teams independently assessed each study and answered all applicable questions. One individual from each pair then reconciled any discrepancies with input from the second person. Study authors of a publication under review were recused from the evaluation of that study. Elements were rated as definitely low RoB, probably low RoB, probably high RoB, definitely high RoB, or not reported (NR). A rating of NR was considered equivalent to probably high-risk RoB.

Data analysis of the human studies

The body of evidence was synthesized qualitatively and, where appropriate, a meta-analysis was performed. In some cases multiple AGD measurements were collected, in these cases, anoscrotal measurements, AGD(as), was preferred in the analysis over anopenile measurements, AGD (ap), because AGD(as) is a more reliable measurement (Sathyanarayana 2015). Phthalate exposure measurements performed in the first trimester were preferred over the second trimester, which was preferred over the third trimester, because the male programming window is approximately gestation weeks 8–14 in the human (Welsh et al. 2008). The sum of DEHP metabolites was the preferred exposure metric over MEHP, which was preferred over any of the other DEHP metabolites, because the sum better reflects the parent compound exposure. For the studies by Bustamonte-Montes et al. (2013) and Swan (2008), the confidence intervals were estimated using the reported p-value, assuming a normal distribution. For other studies, confidence intervals were included in the published manuscript. Slopes (beta coefficients) were reported in the evaluated human literature as units of change in mm per $\log_{10}$ change in urinary concentrations of DEHP metabolites. To standardize effect sizes across studies, each reported beta coefficient was divided by the mean value of the reported outcome measure prior to conducting the meta-analysis. The result is that each beta coefficient was standardized to a percent change in AGD per $\log_{10}$ change in urinary DEHP metabolite concentrations.

Data analysis of the animal studies

The body of evidence was synthesized qualitatively and, where appropriate, a meta-analysis was performed. Summaries of main characteristics for each study used in the meta-analysis were evaluated to determine comparability between studies, identify data transformations necessary to ensure comparability, and determine whether heterogeneity was a concern. The main characteristics considered across all eligible animal studies included: experimental design, species and age, exposure (developmental stage, dosing), health outcomes, type of data and statistical methods, and variation

in the degree of RoB at the individual study level. Animal studies in which exposures did not cover the entire male programming window (gestation days 16–18 in the rat and gestation days 14–16 in the mouse, respectively (Welsh et al. 2008)) were excluded from the meta-analysis. When multiple AGD measures were reported in an animal study the following priority was used: (a) the earliest postnatal time point was used when AGD was measured at multiple time points; and (b) for studies that reported AGD in multiple units, AGD in mm/cube root of body weight was preferred followed by AGD in mm/body weight and AGD in mm. Effect sizes were calculated as follows:

$$y_i = 100 \times In\frac{mean\ of\ treatment\ group\ i}{mean\ of\ concurrent\ control\ group} = 100 \times In\left(1 + \frac{\%change_i}{100}\right)$$

For instance, a −5% change corresponds to y = -5.1. This transformation resulted in confidence intervals that are more symmetric and closer to normal (Lajeunesse 2011). A standard random effects model was applied in the meta-analysis, using the Restricted Maximum Likelihood Estimate as implemented in the R package metafor (Raudenbush 2009; Viechtbauer 2005, 2010).

$$y_i = \mu + u_i + \varepsilon_i$$

where y_i is the observed effect size for group i; μ is the average true effect size; $\mu + u_i$ is true effect size for group i, which is normally distributed $\mu_i \sim N(0,tau^2)$; and $\varepsilon_i \sim N(0, v_i)$ is the sampling error, where v_i is calculated based on the reported sample sizes and standard deviations of the treatment and control groups. In this model, different treatment groups in the same study are treated as independent, even though they usually share a common control group, leading to inter-group correlations. To check the impact of these correlations, one of the sensitivity analyses involved choosing the single highest treatment group from each study, so that each y_i represents a separate study, and is therefore independent. A separate sensitivity analysis involved leaving one study out at a time, to check if any single study was highly influential. The average true effect μ was estimated along with its 95% confidence interval (CI) and z-score. For evaluating heterogeneity, tau^2 was estimated, as well as the Q statistic and its p-value (whether there is statistically significant heterogeneity) and the I^2 index (I^2 = tau^2/overall variance). Rat and mouse data were analyzed separately, due to known anticipated species differences in sensitivity to phthalates (Johnson, Heger, and Boekelheide 2012). In addition, rat data were subjected to a subgroup analysis by strain because of anticipated differential sensitivity across strains (Wilson et al. 2007). Meta-regression of the animal AGD data was also used for benchmark dose (BMD) estimates. Meta-regression involved adding n_j predictors $x_{j,i}$ to the random effects model in an attempt to explain the residual heterogeneity.

$$y_i = \mu + u_i + \sum\nolimits_{j=1\ldots n_j} \beta_j x_{j,i} + \varepsilon_i$$

The meta-regression analyses focused on the dose–response relationship. Three models were used: linear: $x_{1,I}$ = $dose_i$ log-linear: $x_{1,I}$ = $\log_{10}(dose_i)$, and linear-quadratic: $x_{1,I}$ = $dose_i$; $x_{2,I}$ = $dose^2_i$. For the linear and linear-quadratic models, the "intercept" term μ was omitted to ensure that there was no effect at dose = 0. These two models were also used to estimate BMD values based on the average true effect across studies $y_{avg}(dose) = \beta 1 \times dose$ and $y_{avg}(dose) = \beta_1 \times dose + \beta_2 \times dose^2$. As with the standard BMD methodology, Akaike's Information Criterion (AIC) was used to select the preferred model. Covariates such as species and strain were assessed by sub-group analyses.

Confidence rating and level of evidence conclusions

The confidence in each body of evidence was evaluated separately for the human and animal data using the GRADE system for evidence assessment as adapted in the OHAT methodology (Guyatt et al. 2011; Rooney et al. 2014). In brief, data were initially grouped within outcomes by key study design features, and each grouping of investigations was given an initial confidence rating of high, moderate, low, or very low based upon these features. Several factors were then considered to determine whether the initial rating should be downgraded or upgraded (see Table 1). Confidence ratings were independently assessed by two individuals, and discrepancies were

Table 1. Profile of the confidence in the body of evidence on DEHP and AGD in humans and animals. Human studies evaluated DEHP metabolites (metabolites including MEHP; 5-oxo-MEHP; 5OH-MEHP; sumDEHP metabolites). Key: — if no concern or not present; ⊠ if serious concern to downgrade confidence; ⊠ if sufficient to upgrade confidence.

			Factors Decreasing Confidence				Factors Increasing Confidence					
Evidence source	Initial confidence rating (# of studies)	Risk of bias	Unexplained inconsistency	Indirectness	Imprecision	Publication bias	Large magnitude	Dose response	Residual confounding	Consistency across species/models	Rare outcome	**Final confidence rating**
Humans	Moderate (6 prospective)[a]	—	—	—	—	—	—	—	—	—	—	**Moderate**
Animals	High (16 rat,[b] 3 mouse[c])	↓	—	—	—	—	↑	↑	—	—	—	**High**

[a]Swan (2008); Bustamante-Montes et al. (2013); Bornehag et al. (2015); Swan et al. (2015); Jensen et al. (2016); Martino-Andrade et al. (2016).
[b]Moore et al. (2001); Borch et al. (2004); Jarfelt et al. (2005); Wolfe and Layton (2005); Andrade et al. (2006); Culty et al. (2008); Lin et al. (2008), Lin et al. (2009); Christiansen et al. (2009), Christiansen et al. (2010); Gray et al. (2009); Martino-Andrade et al. (2009); Vo et al. (2009); Li et al. (2013); Zhang et al. (2013); Jones et al. (2015).
[c]Liu et al. (2008); Do et al. (2012); Pocar et al. (2012).

resolved by consensus through discussion with a third individual. After a final confidence rating was determined, the rating was translated into a level of evidence using the scheme developed by OHAT (NTP 2015).

Integration of evidence and drawing hazard identification conclusions

The OHAT framework was used to draw hazard identification conclusions (NTP 2015). The procedure involves integrating the levels of evidence ratings for the human and animal data and considering them within the context of biological plausibility provided by mechanistic information. The 5 possible hazard conclusions considered were: (1) known, (2) presumed, (3) suspected, (4) not classifiable, or (5) not identified to be a hazard to humans. If either the animal or the human evidence stream was described as having inadequate evidence, conclusions were drawn based on a single evidence stream.

Results

Search results

Sixteen studies assessing the effect of *in utero* exposure to phthalates on male reproductive effects in humans were identified (Figure 1(a)), and 13 met the inclusion screening criteria. Three studies (Adibi et al. 2015; Barrett et al. 2016; Martino-Andrade et al. 2016) involved "subanalyses" of a cohort by Swan et al. (2015) and one study (Swan 2008) had expanded results from an earlier study on the same cohort by Swan et al. (2005) and included a larger sample size. To avoid double-counting data from the same cohort, the studies from Adibi et al. (2015), Barrett et al. (2016), and Swan et al. (2005) were excluded from data extraction. Martino-Andrade et al. (2016) was retained because it provided additional information beyond Swan et al. (2015) on windows of exposure during the second and third trimester. Of the 13 phthalate studies, 6 studies evaluated DEHP (Figure 1(a) and Table 2).

Seventy studies assessed the effect of *in utero* exposure to phthalates on male reproductive effects in male non-human mammals (Figure 1(b)). Of the 70 studies, 19 studies evaluated DEHP and AGD; 16 studies used the rat model and 3 studies used the mouse model (see Table 3).

RoB evaluation

Most of the human studies had either a low or a very low RoB rating across domains (Figure 2(a)). Although most animal studies had low or very low

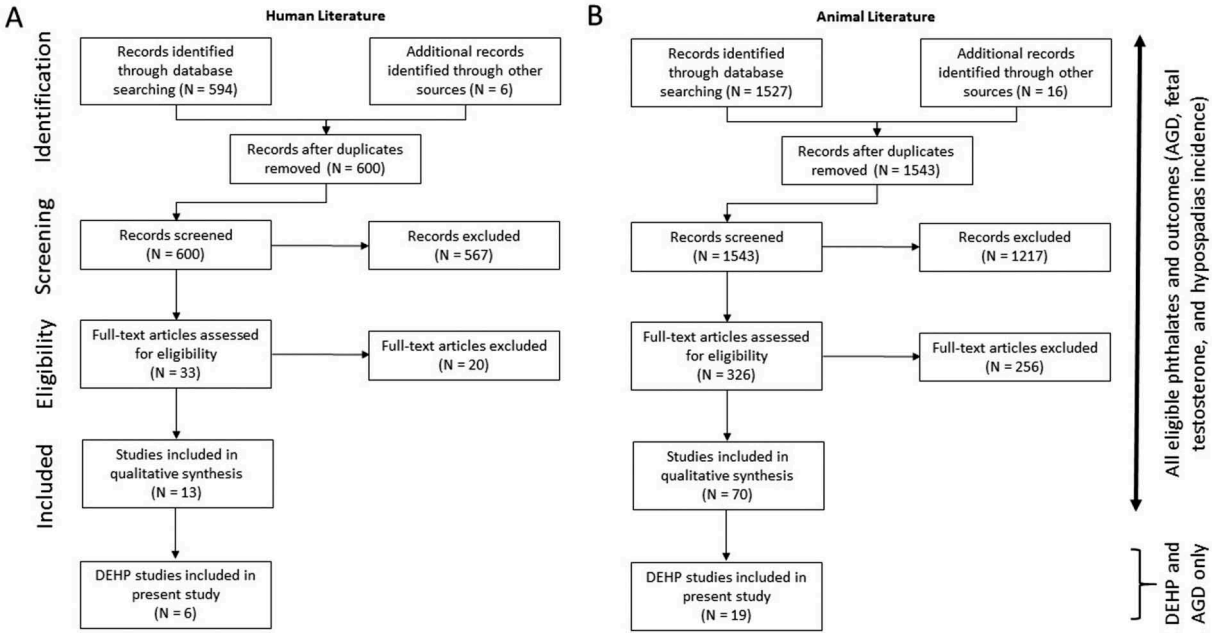

Figure 1. Summary of the search and screening of the literature on the effects of *in utero* exposure to phthalates on male AGD in humans (A) and animals (B). The initial search and screening process identified studies of any ortho phthalate and effects on AGD, fetal testosterone, or hypospadias. The studies were subsequently divided by individual phthalates and by effect. This figure depicts the number of relevant studies on DEHP and effects on AGD.

Table 2. Summary of human studies of DEHP and AGD.

Reference	Study Design, Location, Years, Sample Size	Metabolites Measured in Maternal UrineTime of Measurement, Measurement Value.	Outcome Measures, Age at Measurement	Analytic Method	Confounders Considered
Bornehag et al. (2015)	Prospective cohort, Sweden, 2009–2010, **196**	**sumDEHP metabolites**; MEHP; 5OH-MEHP; 5oxo-MEHP; gestational weeks 9–11. Geometric mean for sumDEHP metabolites = 142.61 nmole/ml.	**AGD (as)** AGD (ap) 21 months	Linear regression of log_{10}-transformed metabolite concentrations	Infant age, weight, gestational week of urine sample, and urinary creatinine
Bustamante-Montes et al. (2013)	Prospective cohort, Mexico (years not specified), **73**	**MEHP; last trimester**. Mean for MEHP = 4 ng/ml.	**AGD (as)** AGD (penis posterior) AGD (ap) 24–48 h	Linear regression of mean exposure	Infant length and urinary creatinine
Jensen et al. (2016)	Prospective cohort, Denmark, 2010–2012, **245 (AGD [as])** 236 (AGD [ap])	**sumDEHP metabolites; gestational week 28**. Mean for sumDEHP metabolites = 29.3 ng/ml	**AGD (as)** AGD (ap) 3 months	Multivariable linear regression of ln-transformed or quartiles metabolite concentrations	Infant age and weight
Suzuki et al. (2012)	Prospective cohort, Japan, 1999–2002, 111	MMP, MEP, MnBP, MBzP, MEHP, MEHHP, MEOHP; mean of 29 ± 9 gestational weeks. Median for MEHP = 3.71 ng/mL.	AGD (as) AGD (ap) At delivery AGD index reported	Multiple regression analysis for each phthalate metabolite or the sum of several phthalate metabolites	Maternal age, smoking status, urinary daidzein and equol concentrations, gestational week, and birth order
Swan (2008)	Prospective cohort, U.S., 1999–2002, **106**	**sumDEHP metabolites**; MEHP; 5OH-MEHP; 5oxo-MEHP; mean of 29 gestational weeks	**AGD (ap)** 13 months (mean)	Regression of log_{10} urinary metabolite concentrations	Infant age and weight
Swan et al. (2015); Martino-Andrade et al. (2016)	Prospective cohort, U.S., 2010–2012, **366 (1st trimester)** 168 (2nd and 3rd trimesters)	**sumDEHP metabolites**; MEHP; 5OH-MEHP; 5oxo-MEHP; 5carboxy-MEPP; all 3 trimesters. Geometric mean for sumDEHP metabolites = 71.7 ng/ml.	**AGD (as)** AGD (ap) At birth or soon thereafter	Regression of log_{10} urinary metabolite concentrations	Infant age, gestational age, maternal age, weight-for-length z-score, time of day of urine collection, maternal age, and study center

NOTE: AGD (ap), distance between the anus and base of the phallus; AGD (as), distance between the anus and scrotum. Studies and outcomes included in meta-analyses are shown in **bold**.

RoB for a majority of domains, several domains were commonly rated as "not reported" because reporting of methods and results was incomplete. Most animal studies were rated as having a probably high RoB (Figure 2(b)) for inadequately describing the method of AGD measurement and/or the reliability of the test methods used to measure AGD (e.g., use of micrometer caliper or reticule micrometer), and not reporting whether blinding of the outcome assessor was performed. In addition, in 8 animal studies, the experimental design and/or statistical methods did not explicitly account for litter effects. There was no qualitative evidence of publication bias in either the animal or human literature.

Effects of DEHP on AGD in humans

Five prospective cohorts contributed data to the analysis of human data on DEHP and AGD (Bornehag et al. 2015; Bustamante-Montes et al. 2013; Jensen et al. 2016; Swan 2008; Swan et al. 2015). Suzuki et al. (2012) was not included in the meta-analysis because it was the only study that reported results only as AGD index (AGD divided by body weight). The results of the meta-analysis are presented in Figure 3 (upper half). In the primary analysis, the data from 5 studies were expressed as beta coefficients standardized to a percent change per log_{10} change in urinary DEHP metabolite concentrations, and analyzed using a random effects model. A significant summary estimate of −4.07 (95% CI: −6.49, −1.66; [p = 0.0009]) was found for the change in AGD per log_{10} rise in urinary DEHP metabolite concentrations. There was no significant heterogeneity, with an estimated I^2 value of 0% (Q statistic was not statistically significant). Two studies (Swan 2008; Swan et al. 2015) accounted for over 60% of the weight in the summary estimate. Figure 3 (bottom half) shows the sensitivity

Table 3. Summary of animal studies of DEHP and AGD. Animals were exposed by gavage unless otherwise indicated.

Reference	Species/Strain	Doses, mg/kg-day	Window of *In Utero* Exposure	AGD Measurement	Age at AGD Measurement
Andrade et al. (2006)	Wistar rat	0, 0.015, 0.045, 0.135, 0.405, 1.215, 5, 15, 45, 135, 405	GD 6-PND 21[d]	AGD (mm)	PND 22
Borch et al. (2004)	Wistar rat	0, 750	GD 7-21[e]	AGD (mm)	PND 3
Christiansen et al. (2009)	Wistar rat	0, 3, 15, 30	GD 7-21[e]	AGD (mm/cube root BW)	PND 0
Christiansen et al. (2010)	Wistar rat:	0, 10, 30, 100, 300, 600, 900	GD 7-PND 16[e]	AGD (mm)	PND 1
	Wistar rat:	0, 3, 10, 30, 100	GD 7-PND 16[e]	AGD (mm)	PND 1
Culty et al. (2008)	Sprague-Dawley rat	0, 234, 469, 700, 750, 938, 1,250	GD 14-PND 0[f]	AGD (mm)	PND 60
Do et al. (2012)	CD-1 mouse	0, 0.0005, 0.001, 0.005, 0.5, 50, 500[a]	GD 9-18[d]	AGD (mm), AGD (mm/g)	GD 18
Gray et al. (2009)	Sprague-Dawley rat	0, 11, 33, 100, 300	GD 8-PND 17[e]	AGD (mm)	PND 2
Jarfelt et al. (2005)	Wistar rat	0, 300, 750	GD 7-PND 3[f]	AGD (mm)	PND 3
Jones et al. (2015)	Sprague-Dawley rat	0, 10	GD 14-21[f]	AGD (mm/g)	PND 3 and 6
Li et al. (2013)	Sprague-Dawley rat	0, 500, 750, 1,000	GD 12-19[d]	AGD (mm), AGD (mm/g)	PND 1
Lin et al. (2008)	Long-Evans rat	0, 10, 100, 750	GD 2-20[f]	AGD (mm)	GD 21
Lin et al. (2009)	Long-Evans rat	0, 10, 750	GD 12.5-PND 21[f]	AGD (mm)	PND 2
Liu et al. (2008)	C57BL/6 mouse	0, 100, 200, 500	GD 12-17[g]	AGD (mm)	GD 19
Martino-Andrade et al. (2009)	Wistar rat	0, 150	GD 13-21[d]	AGD (mm), AGD (mm/cube root BW)	GD 21
Moore et al. (2001)	Sprague-Dawley rat	0, 375, 750, 1,500[b]	GD 13-21[d]	AGD (mm)	PND 1
Pocar et al. (2012)	CD-1 mouse	0, 0.05, 5, 500[c]	GD 0.5-PND 21[g]	AGD (mm/cube root BW)	PND 42
Vo et al. (2009)	Sprague-Dawley rat	0, 10, 100, 500	GD 11-21[d]	AGD (mm)	PND 63
Wolfe and Layton (2005)	Sprague-Dawley rat	0, 321.42, 643.95 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat	0.12, 0.78, 2.37, 7.91, 23.3, 77.45, 592.3, 774.65, 0.12 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat	0.09, 0.48, 1.4, 4.9, 14, 48, 391, 543, 0.09 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat	0.1, 0.47, 1.4, 4.8, 14, 46, 359 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F2 offspring of treated dams)	0.09, 543 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F2 offspring of treated sires)	0.09, 543 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F3 offspring of treated dams)	0.1, 359 [c]	GD 0-parturition	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F3 offspring of treated sires)	0.1, 359 [c]	GD 0-parturition[f]	AGD (mm), AGD (mm/g)	PND 1
Zhang et al. (2013)	Sprague-Dawley rat	0, 250	GD 3-PND 21[d]	AGD (mm/cube root BW)	PND 1 and 22

[a]DEHP administered by micropipetter
[b]Method of administration is unknown
[c]DEHP administered in diet
[d]GD 0 = sperm positive/vaginal plug positive
[e]GD 1 = day after mating
[f]GD 0 = ND
[g]GD 0.5 = sperm positive/vaginal plug positive
NOTE: BW, body weight; GD, gestation day; PND, postnatal day, ND, not defined

A

RoB Domain	Bornehag et al. 2015	Bustamante-Montes et al. 2013	Jensen et al. 2016	Suzuki et al. 2012	Swan 2008	Swan et al. 2015
Appropriate comparison group	NR	++	++	+	+	++
Confounding and modifying variables were accounted for	++	+	+	+	+	++
High attrition/data exclusion	+	+	++	+	+	++
Exposure characterization	++	+	++	+	+	++
Confidence in outcome assessment	++	+	++	-	+	++
All outcomes reported	++	+	+	+	+	++
Other threats to internal validity	+	+	+	+	+	++

Not applicable	NA
Definitely high RoB	-
Probably high RoB	-
Not reported	NR
Probably low RoB	+
Definitely low RoB	++

B

RoB Domain	Rat																Mouse		
	Andrade et al. 2006	Borch et al. 2004	Christiansen et al. 2009	Christiansen et al. 2010	Culty et al. 2008	Gray et al. 2009	Larsen et al. 2005	Jones et al. 2015	Li et al. 2013	Lin et al. 2008	Lin et al. 2009	Martino Andrade et al. 2001	Moore et al. 2001	Wo et al. 2009	Wolfe and Layton 2005	Zhang et al. 2013	Do et al. 2012	Liu et al. 2008	Poon et al. 2012
Exposure adequately randomized	NR	+	+	+	NR	++	+	NR	+	NR	NR	NR	+	NR	++	+	+	+	+
Allocation to groups was concealed	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Identical experimental conditions	+	+	+	+	-	++	+	+	+	+	+	-	-	+	+	+	+	+	-
Blinding of research personnel	+	NR	+	+	NR	+	NR	NR	NR	NR	NR	++	NR	+	NR	NR	NR	NR	NR
High attrition/data exclusion	NR	+	+	+	+	++	-	NR	+	NR	NR	-	+	+	+	+	+	NR	-
Exposure characterization	+	+	+	+	+	+	+	NR	+	NR	NR	+	+	NR	++	NR	-	+	NR
Confidence in outcome assessment	+	+	+	+	-	+	NR	NR	NR	NR	NR	++	NR	NR	NR	++	++	-	NR
All outcomes reported	++	+	-	++	++	++	+	+	++	+	++	++	+	++	++	+	+	NR	-
Other threats to internal validity	++	+	+	++	+	++	+	-	+	NR	+	+	+	-	+	+	NA	+	+
Control for litter effects	++	++	++	++	++	++	++	NR	-	-	NR	++	+	-	NR	++	++	-	NR

Figure 2. Risk of bias heatmap of studies of DEHP and AGD in humans (A) and animals (B). The study by Martino-Andrade et al. (2016) does not appear in the heatmap of the human studies because it is linked to the Swan et al. (2015) study; it has the same risk of bias evaluation as that study. In HAWC: https://hawcproject.org/summary/visual/341/.

analyses that were performed by leaving one study out at a time. Leaving one study out at a time, the summary estimates ranged from –4.35 to –3.59. The summary estimate remained significant in all cases, with p-values ranging from 0.0007 to 0.019. There was no observed heterogeneity in any of these cases (I^2 value of 0%). After the Swan studies, the next largest weight in the summary estimate was obtained from Jensen et al. (2016).

Sensitivity analyses were further performed using alternative effect estimates for each study (Supplemental Table 2). The summary estimates ranged from –4.78 to –1.51. In 11 of the 42 alternative analyses, the summary estimates were no longer significant (summary estimates range from –1.51 to –2.69), with p-values ranging from 0.05 to 0.41. All of the non-significant alternative analyses involved replacing the Swan et al. (2015) results using first trimester DEHP metabolite measurements with results from Martino-Andrade et al. (2016) using second trimester or third trimester DEHP metabolite measurements. Each of these analyses also led to greater heterogeneity (I^2 up to 54%, though none were significant). Finally, 8 additional sensitivity analyses were conducted restricting the included results to more homogeneous exposure and/or outcome measures (e.g., using only the sum DEHP metabolite estimates). The resulting summary estimates ranged from

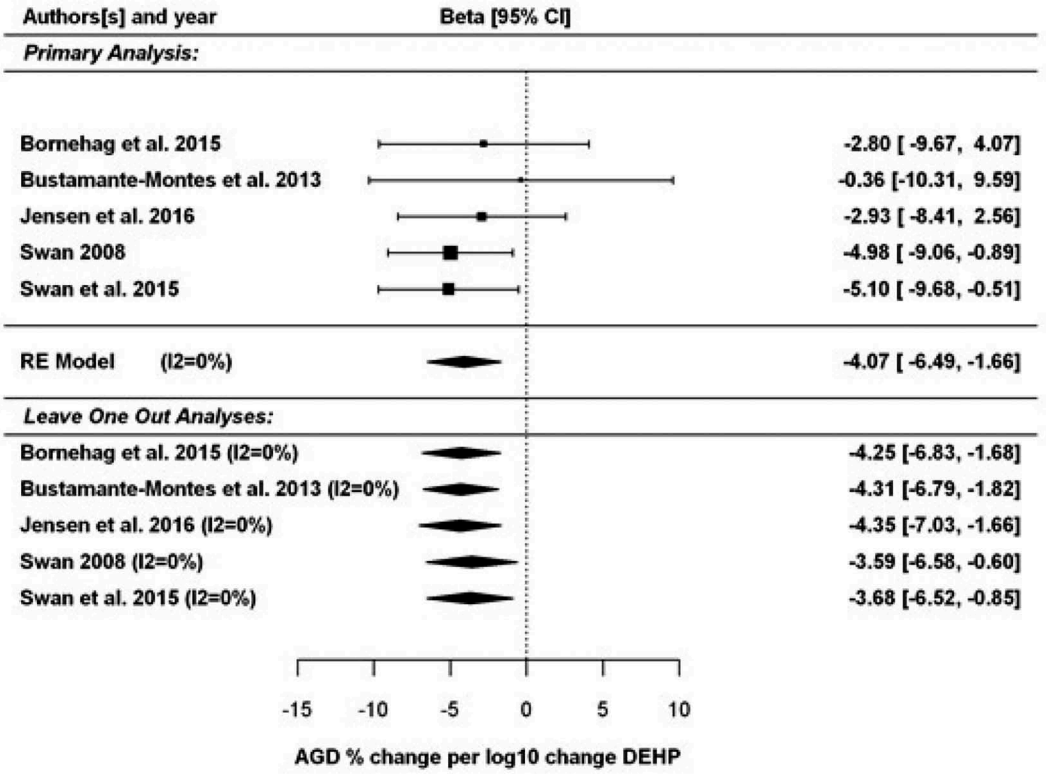

Figure 3. Results of the meta-analysis of studies on DEHP and AGD in humans are shown as the percent change per $\log_{10}$ change in DEHP concentration, including sensitivity analyses leaving one study out at a time. Each box and whisker represents the point estimate and 95% confidence interval for each study, with the size of the box proportional to the inverse-variance weight. Diamonds represent summary estimates under a random effects (RE) model, both overall and with individual studies removed.

−4.2 to −2.0, all of which were significantly different from 0. Further, there was no observed heterogeneity in any of these cases ($I^2 = 0$).

Effects of DEHP on AGD in animals

In all, 13 of the 16 rat studies and all 3 of the mouse studies were included in the analysis; 3 of the rat studies were excluded because they were missing group size values (Borch et al. 2004; Jones et al. 2015; Vo et al. 2009). A significant overall reduction in rat AGD was found (–3.96 [95% confidence interval (CI)]: –5.07, –2.85). Significant linear trends in $\log_{10}$(dose) (–1.97 [95% CI: –2.98, –0.96]) and dose (–1.55 [95% CI: –1.86, –1.24]) were also noted. No single study was found to influence the results, as the overall effect was robust to leaving out individual studies. Using the linear-quadratic model, there was low heterogeneity (I^2 = 23%, p = 0.12), with a BMD_5 estimated to be 270 mg/kg-day (95% CI: 180, 420). When analysis was restricted to the highest dose group, there was a larger overall effect, larger linear trend in $\log_{10}$(dose), consistent linear trend in dose, and consistent BMD_5 estimates. In subgroup analyses, there were significant overall effects and linear trends in $\log_{10}$(dose) and dose for Sprague-Dawley and Wistar rats separately, with reduced heterogeneity. Sprague-Dawley rats appeared somewhat less sensitive than Wister rats, with smaller overall effect sizes, smaller trend in $\log_{10}$(dose), and larger benchmark dose estimates. Specifically, a BMD_5 for Sprague-Dawley rats was estimated to be 290 mg/kg-day (95% CI: 170, >1,000), whereas the BMD_5 for Wistar rats was estimated to be 150 mg/kg-day (95% CI: 100, 280). The results of linear-quadratic meta-regression, the model with the lowest Akaike information criterion (AIC) corrected for finite sample sizes (AIC_c), are shown in Figure 4.

No significant overall effect for DEHP and changes in mouse AGD were seen, but there were significant linear trends in $\log_{10}$(dose) (–1.77 [95% CI: –2.71, –0.85]) and dose (–2.03 [95% CI: –3.51, –0.55]). Under

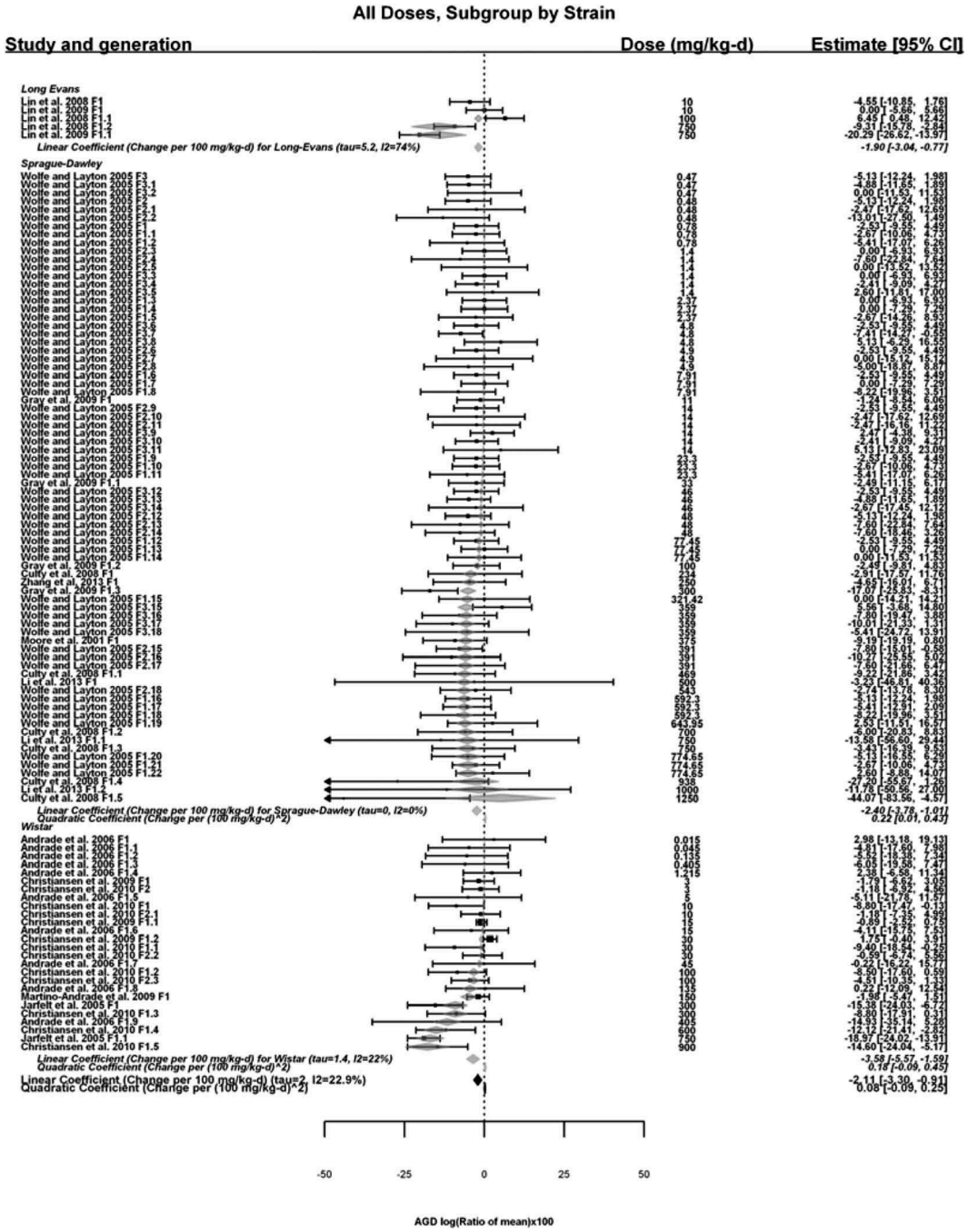

Figure 4. Results of the meta-regressions of studies on DEHP and AGD in rats. The overall effect of treatment in each strain is shown at the bottom of each subgroup analysis above as the change per 100 mg/kg-day. Each box and whisker represents the point estimate and 95% confidence interval for each study, with the size of the box proportional to the inverse-variance weight. Grey diamonds that overlap with the box and which represents the predicted effect size (95% CI) under the meta-regression; grey diamonds alone represent the fitted linear and/or quadratic coefficients (95% CI) for the subgroup; and black diamonds alone represent the fitted coefficients (95% CI) overall.

the linear-quadratic model, there was low heterogeneity (0%, p = 0.19), with the BMD_5 estimated to be 110 mg/kg-day (95% CI: 90, 150). When the analysis was restricted to the highest dose group, there remained no significant overall effect, and there was no longer a significant linear trend. The overall effect was no longer significant when leaving out some individual studies during the sensitivity analyses. The results for the overall effect estimate, which had the lowest AICc, are illustrated in Figure 5.

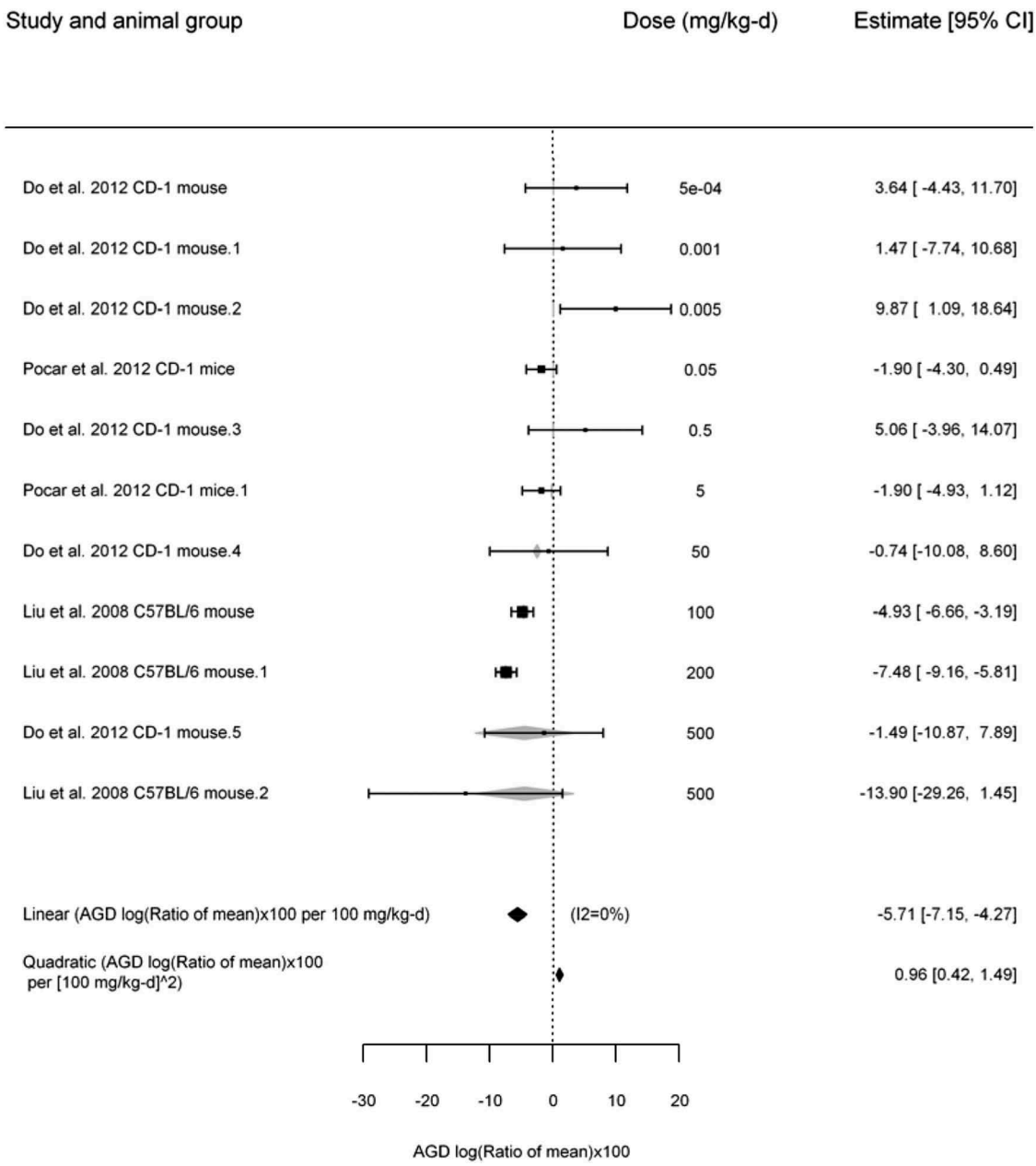

Figure 5. Results of the meta-analysis of studies on DEHP and AGD in mice. The overall effect of treatment is shown at the bottom of the figure as the change per 100 mg/kg-day. Each box and whisker represents the point estimate and 95% confidence interval for each study, with the size of the box proportional to the inverse-variance weight. Grey diamonds that overlap with the box and which represents the predicted effect size (95% CI) under the meta-regression; and black diamonds alone represent the fitted linear and quadratic coefficients (95% CI) overall.

Confidence in the body of evidence

The human studies involved exposures that occurred prior to the outcome; outcomes were measured on individuals; and the inclusion of a (control) comparison group. These factors led to an initial moderate rating for the confidence in the human studies (Table 1). There were no marked changes in the confidence rating for the human evidence after considering factors that might increase or decrease confidence.

The initial rating for the confidence in the animal studies was high (Table 1) because they involved controlled exposures, exposures occurred prior to the outcome, outcomes were measured on individual animals, and a concurrent control comparison group was used. One factor that decreased

confidence was the concern of significant RoB (described above under "RoB Evaluation") related to confidence in the reliability of outcome measure, blinding of investigators to the treatment groups, and control for litter effects. However, confidence was not downgraded because two factors increased confidence in the evidence, a large magnitude of effect and evidence of a dose-response relationship.

Level of evidence for the health effect

The results from the human studies demonstrate a consistent pattern of findings that higher maternal urinary concentrations of DEHP metabolites during the prenatal male genital programming window are associated with a smaller AGD in male infants compared to infants whose mothers contained lower DEHP exposures during pregnancy. In accordance with the OHAT method (NTP, 2015), evidence of an effect combined with a moderate confidence rating in the body of evidence leads to the conclusion that there is a moderate level of evidence in humans that fetal exposure to DEHP is associated with a reduction in AGD.

A meta-analysis of the animal studies found consistent evidence of a decrease in AGD after *in utero* exposure to DEHP in rats only; thus, rat data was used in the final evidence integration step. Evidence of an effect and a high confidence rating in the body of evidence led to the conclusion that there is a high level of evidence in male rats that fetal exposure to DEHP is associated with a reduction in AGD.

Evidence integration

The human and animal bodies of evidence present a consistent pattern of findings that fetal exposure to DEHP is associated with reduced AGD in male offspring. Changes in AGD are considered an adverse effect. Under the OHAT method, a combination of a moderate level of human evidence and a high level of animal evidence leads to the conclusion that DEHP *is presumed to be a reproductive hazard to humans on the basis of effects on AGD.*

Discussion

This study found consistent moderate evidence that increasing maternal prenatal urinary concentrations of the sum of DEHP metabolites was associated with reduced AGD in male infants. The present study also found consistent high evidence of decreased AGD in male rats after fetal exposure to DEHP, with a modest dose-response gradient. Integration of these results supported the conclusion that DEHP is presumed to be a reproductive hazard to humans on the basis of effects on AGD. This conclusion is further supported by a recent NASEM report that included the result of animal and human SRs that also considered changes in fetal testosterone levels and hypospadias incidence following *in utero* DEHP exposure (NASEM 2017). Unlike the present study, NASEM conclusions concerning DEHP effects on fetal testosterone ("presumed to be a reproductive hazard to humans') and hypospadias ("suspected to be a reproductive hazard to humans") rested solely on animal evidence since insufficient human evidence was available to assess whether exposure to DEHP is associated with these outcomes (NASEM 2017). Our conclusion that DEHP is presumed to be a reproductive hazard to humans is also supported by mechanistic evidence, including an adverse outcome pathway (AOP), from the rat of reproductive toxicity after *in utero* phthalate exposure during the period of sexual differentiation (Howdeshell et al. 2015). Maternal rat exposure to endocrine-active phthalates during late gestation might reduce the expression of genes encoding proteins involved in steroidogenesis in the fetal testis Leydig cells. Affected proteins can include CYP11A1, CYP17A1, translocator protein (18-kDa) (TSPO), and steroidogenic acute regulatory protein (STAR) (Borch et al. 2006; Gray et al. 2000). Decreased steroidogenic protein expression in fetal rat Leydig cells can result in diminished fetal testis testosterone production. If exposure occurs during the *in utero* male programming window, developmental alterations in androgen-dependent tissues might occur, including reduced expansion of the perineum resulting in a decrease in AGD and altered urethral epithelium closure resulting in hypospadias (Howdeshell et al. 2015). However, the molecular initiating event producing reductions in fetal testis

steroidogenic mRNA levels remains unknown. Data from xenograft studies in which fetal rat, mouse, and human testes were implanted in nude rats or mice exposed to dibutyl phthalate (DBP) indicate that human testes appear to be less sensitive to the effects of phthalates on steroidogenic gene expression when compared with the rat (Heger et al. 2012; Mitchell et al. 2012).

Some animal strain and species differences were seen in the present study. For example, Sprague-Dawley rats were less sensitive than Wistar rats to DEHP effects on AGD, with a BMD_5 of around 300 mg/kg-day compared to 150 mg/kg-day, whereas mice have a BMD_5 of 250–350 mg/kg-day. These strain and species differences may relate to DEHP effect on fetal testosterone production. Studies demonstrated that reproductive-tract malformations were found in male rats when fetal testosterone production was reduced by about 25–70% (Howdeshell et al. 2015). The association between decreases in fetal testosterone and changes in AGD in other species is less clear. For example, studies conducted in mice with a structurally related phthalate, di-n-butyl phthalate (DBP), noted that reduced fetal testicular testosterone occurs in rats, but not in mice, following *in utero* exposure (Gaido et al. 2007; Johnson, Heger, and Boekelheide 2012). In the present study, a significant overall effect on AGD and DEHP exposure was not seen in male mice, thus rat data were relied upon in the final analysis.

When the human and experimental animal data were analyzed similarly (estimating the magnitude of change for each $\log_{10}$ increase in DEHP), the effect estimates were similar: 2–6% for humans, 0–2% for Sprague–Dawley rats, 1–5% for Wistar rats, and 1–3% for mice. Thus, these estimates were largely concordant; however, the dose ranges in which these estimates are observed differ substantially between humans and rats. For example, estimates of mean daily intake in adults in the US population range from approximately 0.0006–0.002 mg/kg-day (Lorber, Angerer, and Koch 2010) to 0.011 mg/kg-day (Lorber and Calafat 2012). Published estimates of daily DEHP intake in adults were therefore several 1000-fold lower than the calculated BMD_5 (>100 mg/kg-day) for effects of DEHP on AGD in rodents. However, intake is not necessarily reflective of biologically active dose, due to potential differences in pharmacokinetics. Fetal DEHP metabolite concentrations have not been measured in human studies but when one considers urinary and amniotic fluid MEHP concentrations, differences between biomarkers of exposure in humans versus animals dosed near the BMD_5 were substantially reduced as compared to the daily intake estimates. For instance, one human study reported a median amniotic fluid MEHP concentration of 23 ng/ml (Huang et al. 2009), which is only 3-fold lower than the mean amniotic fluid MEHP concentration of 68 ng/ml reported in pregnant rats exposed to 11 mg/kg-d (Calafat et al. 2006). In addition, peak concentrations and area under the curves for MEHP and DEHP in human serum in human volunteers given deuterated DEHP are markedly greater than those noted for either rats or marmosets given comparable administered doses (Kessler et al. 2012; Koch, Bolt, and Angerer 2004).

In humans, several studies reported that newborns with hypospadias or cryptorchidism exhibited shorter AGD than infants without these abnormalities (Hsieh et al. 2012; Jain and Singal 2013; Thankamony et al. 2014). In addition, several cross-sectional studies in adult males found that men who have reduced fertility, including lower sperm concentration, count, and motility, display shortened AGD (Eisenberg et al. 2011, 2012; Mendiola et al. 2011; Eisenberg and Lipshultz 2015). However, the quantitative relationship between reduced AGD and increased risk of apical endpoints remains uncertain.

Meta-analysis of animal data remains in its early stages and this study provides a novel illustration of how meta-analysis and meta-regression might be used to integrate animal data across studies. Meta-analysis techniques are more robust than relying on individual study results, because they (1) account for potential heterogeneity across studies, (2) consider the number of studies when calculating variance and 95% confidence intervals for meta-estimates, and (3) increase statistical power as compared to each individual study. Further, results of the meta-analyses of animal and human evidence directly contributed to consideration of the "down/upgrading" factors of imprecision (through summary estimates),

unexplained inconsistency (through analyses of heterogeneity and subgroup analyses), large magnitude of effect (through summary estimates and meta-regression), and dose-response gradient (through meta-regression). Meta-analyses of the animal evidence also supported the estimate of benchmark doses for DEHP effects on AGD. Further, a rigorous and unbiased inclusion of studies that used a common endpoint measured across the animal and human studies enabled consistent meta-analyses approaches between these two bodies of evidence.

In conclusion, our results show that exposure of the fetus to DEHP is associated with decreased AGD in male offspring and DEHP is presumed to be a reproductive hazard to humans. Future risk assessments for DEHP need to take into account that humans are exposed to multiple anti-androgenic phthalates, some of which may confound associations of DEHP with AGD. Further, exposure to multiple anti-androgenic phthalates necessitates the need for risk assessors to consider cumulative risk from DEHP and other anti-androgenic phthalates (Howdeshell, Hotchkiss, and Gray 2017; NRC 2008). Our study also illustrates how animal and human evidence may be identified, analyzed, and integrated to draw hazard conclusions using a systematic methodologic approach. Meta-analyses of the animal and human evidence helped strengthen this data integration and supported quantitative evaluation of the relationships between endocrine active chemicals and endpoints of interest to inform the confidence ratings of the bodies of evidence.

Acknowledgments

Funding was provided by a contract to the National Academies of Sciences, Engineering, and Medicine from the US Environmental Protection Agency. The authors thank Robyn Blain and Pamela Hartman at ICF International for their assistance with data extraction and Jaime Blanck at the William H. Welch Medical Library at Johns Hopkins University for her assistance with planning and performing the literature searches to support the systematic reviews. Selected tables, figures, and selected text from a 2017 report of the National Academies of Sciences, Engineering, and Medicine, *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals* are reprinted with permission from the National Academies Press, Washington, DC.

Competing financial interest declaration

The authors declare they have no actual or potential competing financial interests.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adibi, J. J., M. K. Lee, A. I. Naimi, E. Barrett, R. H. Nguyen, S. Sathyanarayana, Y. Zhao, M. P. Thiet, J. B. Redmon, and S. H. Swan. 2015. Human chorionic gonadotropin partially mediates phthalate association with male and female anogenital distance. *Journal of Clinical Endocrinology and Metabolism* 100:E1216–E1224. doi:10.1210/jc.2015-2370.

Andrade, A. J., S. W. Grande, C. E. Talsness, K. Grote, A. Golombiewski, A. Sterner-Kock, and I. Chahoud. 2006. A dose-response study following in utero and lactational exposure to di-(2-ethylhexyl) phthalate (DEHP): Effects on androgenic status, developmental landmarks and testicular histology in male offspring rats. *Toxicology* 225:64–74. doi:10.1016/j.tox.2006.05.007.

Barrett, E. S., L. E. Parlett, S. Sathyanarayana, J. B. Redmon, R. H. Nguyen, and S. H. Swan. 2016. Prenatal stress as a modifier of associations between phthalate exposure and reproductive development: Results from a multicentre pregnancy cohort study. *Paediatric and Perinatal Epidemiology* 30:105–14. doi:10.1111/ppe.12264.

Borch, J., O. Ladefoged, U. Hass, and A. M. Vinggaard. 2004. Steroidogenesis in fetal male rats is reduced by DEHP and DINP, but endocrine effects of DEHP are not modulated by DEHA in fetal, prepubertal and adult male rats. *Reproductive Toxicology* 18:53–61. doi:10.1016/j.reprotox.2003.10.011.

Borch, J., S. B. Metzdorff, A. M. Vinggaard, L. Brokken, and M. Dalgaard. 2006. Mechanisms underlying the antiandrogenic effects of diethylhexyl phthalate in fetal rat testis. *Toxicology* 223:144–55. doi:10.1016/j.tox.2006.03.015.

Bornehag, C. G., F. Carlstedt, B. A. Jönsson, C. H. Lindh, T. K. Jensen, A. Bodin, C. Jonsson, S. Janson, and S. H. Swan. 2015. Prenatal phthalate exposures and anogenital distance in Swedish boys. *Environmental Health Perspectives* 123:101–07. doi:10.1289/ehp.1408163.

Bustamante-Montes, L. P., M. A. Hernández-Valero, D. Flores-Pimentel, M. García-Fábila, A. Amaya-Chávez, D. B. Barr, and V. H. Borja-Aburto. 2013. Prenatal exposure to phthalates is associated with decreased anogenital distance and penile size in male newborns. *Journal of Developmental Origins of Health and Disease* 4:300–06. doi:10.1017/S2040174413000172.

Calafat, A. M., J. W. Brock, M. J. Silva, L. E. Gray Jr, J. A. Reidy, D. B. Barr, and L. L. Needham. 2006. Urinary and amniotic fluid levels of phthalate monoesters in rats after the oral administration of di(2-ethylhexyl) phthalate and

di-n-butyl phthalate. *Toxicology* 217:22–30. doi:10.1016/j.tox.2005.08.013.

CDC (Centers for Disease Control and Prevention). 2009. *Fourth report on human exposure to environmental chemicals*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention [online]. Accessed February 28, 2018. http://www.cdc.gov/exposurereport/.

CHAP (Chronic Hazard Advisory Panel). 2014. *Report to the U.S. consumer product safety commission by the chronic hazard advisory panel on phthalates and phthalate alternatives*. Bethesda, MD: U.S. Consumer Product Safety Commission, Directorate for Health Sciences. [online]. Accessed February 28, 2018. https://www.cpsc.gov/PageFiles/169902/CHAP-REPORT-With-Appendices.pdf.

Christiansen, S., J. Boberg, M. Axelstad, M. Dalgaard, A. M. Vinggaard, S. B. Metzdorff, and U. Hass. 2010. Low-dose perinatal exposure to di(2-ethylhexyl) phthalate induces anti-androgenic effects in male rats. *Reproductive Toxicology* 30:313–21. doi:10.1016/j.reprotox.2010.04.005.

Christiansen, S., M. Scholze, M. Dalgaard, A. M. Vinggaard, M. Axelstad, A. Kortenkamp, and U. Hass. 2009. Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. *Environmental Health Perspectives* 117:1839–46. doi:10.1289/ehp.0900689.

Culty, M., R. Thuillier, W. Li, Y. Wang, D. B. Martinez-Arguelles, C. G. Benjamin, K. M. Triantafilou, B. R. Zirkin, and V. Papadopoulos. 2008. In utero exposure to di-(2-ethylhexyl) phthalate exerts both short-term and long-lasting suppressive effects on testosterone production in the rat. *Biology of Reproduction* 78:1018–28. doi:10.1095/biolreprod.107.065649.

Do, R. P., R. W. Stahlhut, D. Ponzi, F. S. Vom Saal, and J. A. Taylor. 2012. Non-monotonic dose effects of in utero exposure to di(2-ethylhexyl) phthalate (DEHP) on testicular and serum testosterone and anogenital distance in male mouse fetuses. *Reproductive Toxicology* 34:614–21. doi:10.1016/j.reprotox.2012.09.006.

Eisenberg, M. L., and L. I. Lipshultz. 2015. Anogenital distance as a measure of human male fertility. *Journal of Assisted Reproduction and Genetics* 32:479–84. doi:10.1007/s10815-014-0410-1.

Eisenberg, M. L., M. Shy, R. C. Walters, and L. I. Lipshultz. 2012. The relationship between anogenital distance and azoospermia in adult men. *International Journal of Andrology* 35:726–30. doi:10.1111/j.1365-2605.2012.01275.x.

Eisenberg, M. L., M. H. Hsieh, R. C. Walters, R. Krasnow, and L. I. Lipshultz. 2011. The relationship between anogenital distance, fatherhood, and fertility in adult men. *PLoS ONE* 6:0018973. doi:10.1371/journal.pone.0018973.

Fennell, T. R., W. L. Krol, S. C. Sumner, and R. W. Snyder. 2004. Pharmacokinetics of dibutyl phthalate in pregnant rats. *Toxicological Sciences* 82:407–18. doi:10.1093/toxsci/kfh294.

Furr, J. R., C. S. Lambright, V. S. Wilson, P. M. Foster, and L. E. Gray Jr. 2014. A short-term in vivo screen using fetal testosterone production, a key event in the phthalate adverse outcome pathway, to predict disruption of sexual differentiation. *Toxicological Sciences* 140:403–24. doi:10.1093/toxsci/kfu081.

Gaido, K. W., J. B. Hensley, D. Liu, D. G. Wallace, S. Borghoff, K. J. Johnson, S. J. Hall, and K. Boekelheide. 2007. Fetal mouse phthalate exposure shows that gonocyte multinucleation is not associated with decreased testicular testosterone. *Toxicological Sciences* 97:491–503. doi:10.1093/toxsci/kfm049.

Gray, L. E.Jr, J. Ostby, J. Furr, M. Price, D. N. Veeramachaneni, and L. Parks. 2000. Perinatal exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. *Toxicological Sciences* 58:350–65.

Gray, L. E.Jr, N. J. Barlow, K. L. Howdeshell, J. S. Ostby, J. R. Furr, and C. L. Gray. 2009. Transgenerational effects of di(2-ethylhexyl) phthalate in the male CRL:CD(SD) rat: Added value of assessing multiple offspring per litter. *Toxicological Sciences* 110:411–25. doi:10.1093/toxsci/kfp109.

Guyatt, G. H., A. D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P. J. Devereaux, V. M. Montori, B. Freyschuss, and G. Vist. 2011. GRADE guidelines 6. Rating the quality of evidence—Imprecision. *Journal of Clinical Epidemiology* 64:1283–93. doi:10.1016/j.jclinepi.2011.01.012.

Heger, N. E., S. J. Hall, M. A. Sandrof, E. V. McDonnell, J. B. Hensley, E. N. McDowell, K. A. Martin, K. W. Gaido, K. J. Johnson, and K. Boekelheide. 2012. Human fetal testis xenografts are resistant to phthalate-induced endocrine disruption. *Environmental Health Perspectives* 120:1137–43. doi:10.1289/ehp.1104711.

Higgins, J. P. T., and S. Green 2011. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Accessed July 10, 2018. http://handbook-5-1.cochrane.org/.

Hooijmans, C. R., A. Tillema, M. Leenaars, and M. Ritskes-Hoitinga. 2010. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. *Laboratory Animals* 44:170–75. doi:10.1258/la.2010.009117.

Howdeshell, K. L., A. K. Hotchkiss, and L. E. Gray Jr. 2017. Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. *International Journal of Hygiene and Environmental Health* 220:179–88. doi:10.1016/j.ijheh.2016.11.007.

Howdeshell, K. L., C. V. Rider, V. S. Wilson, J. R. Furr, C. R. Lambright, and L. E. Gray Jr. 2015. Dose addition models based on biologically relevant reductions in fetal testosterone accurately predict postnatal reproductive tract alterations by a phthalate mixture in rats. *Toxicological Sciences* 148:488–502. doi:10.1093/toxsci/kfv196.

Hsieh, M. H., M. L. Eisenberg, A. B. Hittelman, J. M. Wilson, G. E. Tasian, and L. S. Baskin. 2012. Caucasian male infants and boys with hypospadias exhibit reduced anogenital distance. *Human Reproduction* 27:1577–80. doi:10.1093/humrep/des087.

Huang, P. C., P. L. Kuo, Y. Y Chou, and S. J. Lin SJ, and C. C. Lee. 2009. Association between prenatal exposure to

phthalates and the health of newborns. *Environment International* 35: 14–20. doi:10.1016/j.envint.2008.05.012.

Jain, V. G., and A. K. Singal. 2013. Shorter anogenital distance correlates with undescended testis: A detailed genital anthropometric analysis in human newborns. *Human Reproduction* 28:2343–49. doi:10.1093/humrep/det286.

Jarfelt, K., M. Dalgaard, U. Hass, J. Borch, H. Jacobsen, and O. Ladefoged. 2005. Antiandrogenic effects in male rats perinatally exposed to a mixture of di(2-ethylhexyl) phthalate and di(2-ethylhexyl) adipate. *Reproductive Toxicology* 19:505–15. doi:10.1016/j.reprotox.2004.11.005.

Jensen, T. K., H. Frederiksen, H. B. Kyhl, T. H. Lassen, S. H. Swan, C. G. Bornehag, N. E. Skakkebaek, K. M. Main, D. V. Lind, S. Husby, et al. 2016. Prenatal exposure to phthalates and anogenital distance in male infants from a low-exposed Danish cohort (2010–2012). *Environmental Health Perspectives* 124:1107–13. doi:10.1289/ehp.1509870.

Johns, L. E., G. S. Cooper, A. Galizia, and J. D. Meeker. 2015. Exposure assessment issues in epidemiology studies of phthalates. *Environment International* 85:27–39. doi:10.1016/j.envint.2015.08.005.

Johnson, K. J., N. E. Heger, and K. Boekelheide. 2012. Of mice and men (and rats): Phthalate-induced fetal testis endocrine disruption is species-dependent. *Toxicological Sciences* 129:235–48. doi:10.1093/toxsci/kfs206.

Jones, S., A. Boisvert, S. Francois, L. Zhang, and M. Culty. 2015. In utero exposure to di-(2-ethylhexyl) phthalate induces testicular effects in neonatal rats that are antagonized by genistein cotreatment. *Biology of Reproduction* 93:92. doi:10.1095/biolreprod.115.129098.

Kessler, W., W. Numtip, K. Grote, G. A. Csanády, I. Chahoud, and J. G. Filser. 2004. Blood burden of di(2-ethylhexyl) phthalate and its primary metabolite mono(2-ethylhexyl) phthalate in pregnant and nonpregnant rats and marmosets. *Toxicology and Applied Pharmacology* 195:142–53. doi:10.1016/j.taap.2003.11.014.

Kessler, W., W. Numtip, W. Völkel, E. Seckin, G.A. Csanády, C. Pütz, D. Klein, H. Fromme, and J.G. Filser. 2012. Kinetics of di(2-ethylhexyl) phthalate (dehp) and mono (2-ethylhexyl) phthalate in blood and of dehp metabolites in urine of male volunteers after single ingestion of ring-deuterated DEHP. *Toxicology and Applied Pharmacology* 264 (2):284–91. doi: 10.1016/j.taap.2012.08.009.

Koch, H. M., H. M. Bolt, and J. Angerer. 2004. Di(2-ethylhexyl)phthalate (DEHP) metabolites in human urine and serum after a single oral dose of deuterium-labelled DEHP. *Archives of Toxicology* 78:123–30. doi:10.1007/s00204-003-0522-3.

Lajeunesse, M. J. 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology* 92:2049–55. doi:10.1890/11-0423.1.

Li, M., L. Qiu, Y. Zhang, Y. Hua, S. Tu, Y. He, S. Wen, Q. Wang, and G. Wei. 2013. Dose-related effect by maternal exposure to di-(2-ethylhexyl) phthalate plasticizer on inducing hypospadiac male rats. *Environmental Toxicology and Pharmacology* 35:55–60. doi:10.1016/j.etap.2012.10.006.

Lin, H., Q. Q. Lian, G. X. Hu, Y. Jin, Y. Zhang, D. O. Hardy, C. M. Sottas, X. K. Li, and M. P. Hardy. 2008. Involvement of testicular growth factors in fetal Leydig cell aggregation after exposure to phthalate in utero. *Proceedings of the National Academy of Sciences* 105:7218–22. doi:10.1073/pnas.0709260105.

Lin, H., Q. Q. Lian, G. X. Hu, Y. Jin, Y. Zhang, D. O. Hardy, G. R. Chen, Z. Q. Lu, C. M. Sottas, M. P. Hardy, and R. S. Ge. 2009. In utero and lactational exposures to diethylhexyl-phthalate affect two populations of Leydig cells in male Long-Evans rats. *Biology of Reproduction* 80:882–88. doi:10.1095/biolreprod.108.072975.

Lioy, P. J., R. Hauser, C. Gennings, H. M. Koch, P. E. Mirkes, B. A. Schwetz, and A. Kortenkamp. 2015. Assessment of phthalates/phthalate alternatives in children's toys and childcare articles: Review of the report including conclusions and recommendation of the chronic hazard advisory panel of the consumer product safety commission. *Journal of Exposure Science & Environmental Epidemiology* 5:343–53. doi:10.1038/jes.2015.33.

Liu, X., D. W. He, D. Y. Zhang, T. Lin, and G. H. Wei. 2008. Di (2-ethylhexyl) phthalate (DEHP) increases transforming growth factor-beta1 expression in fetal mouse genital tubercles. *Journal of Toxicology and Environmental Health, Part A* 71:1289–94. doi:10.1080/15287390802114915.

Lorber, M., and A. M. Calafat. 2012. Dose reconstruction of di(2-ethylhexyl) phthalate using a simple pharmacokinetic model. *Environmental Health Perspectives* 120:1705–10. doi:10.1289/ehp.1205182.

Lorber, M., J. Angerer, and H. M. Koch. 2010. A simple pharmacokinetic model to characterize exposure of Americans to di-2-ethylhexyl phthalate. *Journal of Exposure Science and Environmental Epidemiology* 20:38–53. doi:10.1038/jes.2008.74.

Martino-Andrade, A. J., F. Liu, S. Sathyanarayana, E. S. Barrett, J. B. Redmon, R. H. Nguyen, H. Levine, and S. H. Swan; TIDES Study Team. 2016. Timing of prenatal phthalate exposure in relation to genital endpoints in male newborns. *Andrology* 4: 585–93. doi: 10.1111/andr.12180.

Martino-Andrade, A. J., R. N. Morais, G. G. Botelho, G. Muller, S. W. Grande, G. B. Carpentieri, G. M. Leão, and P. R. Dalsenter. 2009. Coadministration of active phthalates results in disruption of foetal testicular function in rats. *International Journal of Andrology* 32:704–12. doi:10.1111/j.1365-2605.2008.00939.x.

Mendiola, J., R. W. Stahlhut, N. Jørgensen, F. Liu, and S. H. Swan. 2011. Shorter anogenital distance predicts poorer semen quality in young men in Rochester, New York. *Environmental Health Perspectives* 119:958–63. doi:10.1289/ehp.1103421.

Mitchell, R. T., A. J. Childs, R. A. Anderson, S. van den Driesche, P. T. Saunders, C. McKinnell, W. H. Wallace, C. J. Kelnar, and R. M. Sharpe. 2012. Do phthalates affect steroidogenesis by the human fetal testis? Exposure of human fetal testis xenografts to di-n-butyl phthalate. *Journal of Clinical Endocrinology and Metabolism* 97: E341–E348. doi:10.1210/jc.2011-2411.

Moore, R. W., T. A. Rudy, T. M. Lin, K. Ko, and R. E. Peterson. 2001. Abnormalities of sexual development in male rats with in utero and lactational exposure to the antiandrogenic plasticizer di(2-ethylhexyl) phthalate. *Environmental Health Perspectives* 109:229–37. PMC1240240.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2017. *Application of systematic review methods in an overall strategy for evaluating low-dose toxicity from endocrine active chemicals*. Washington, DC: The National Academies Press.

Negev, M., T. Berman, S. Reicher, S. Balan, A. Soehl, S. Goulden, R. Ardi, Y. Shammai, L. Hadar, A. Blum, and M. L. Diamond. 2018. Regulation of chemicals in children's products: How U.S. and EU regulation impacts small markets. *Science of the Total Environment* 616-617:462–71. doi:10.1016/j.scitotenv.2017.10.198.

NRC (National Research Council). 2008. *Phthalates and cumulative risk assessment: The tasks ahead*. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015. Handbook for conducting a literature-base health assessment using OHAT approach for systematic review and evidence integration, January 9, 2015. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences [online]. Accessed May 2, 2018. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

Pocar, P., N. Fiandanese, C. Secchi, A. Berrini, B. Fischer, J. S. Schmidt, et al. 2012. Exposure to di(2-ethyl-hexyl) phthalate (DEHP) in utero and during lactation causes long-term pituitary-gonadal axis disruption in male and female mouse offspring. *Endocrinology* 153:937–48. doi:10.1210/en.2011-1450.

Raudenbush, S. W. 2009. Analyzing effect sizes: Random-effects models. In *The handbook of research synthesis and meta-analysis*, eds. H. Cooper, L. V. Hedges, and J. C. Valentine, 295–315. 2nd ed. New York: Russell Sage Foundation.

Rooney, A. A., A. L. Boyles, M. S. Wolfe, J. R. Bucher, and K. A. Thayer. 2014. Systematic review and evidence integration for literature-based environmental health assessments. *Environmental Health Perspectives* 122:711–18. doi:10.1289/ehp.1307972.

Sathyanarayana, S., R. Grady, J. B. Redmon, K. Ivicek, E. Barrett, S. Janssen, R. Nguyen, S. H. Swan, and T. I. D. E. S. Study Team. 2015. Anogenital distance and penile width measurements in The Infant Development and the Environment Study (TIDES): Methods and predictors. *Journal of Pediatric Urology* 11:76.e1–e6. doi:10.1016/j.jpurol.2014.11.018.

Silva, M. J., J. A. Reidy, A. R. Herbert, J. L. Preau Jr, L. L. Needham, and A. M. Calafat. 2004. Detection of phthalate metabolites in human amniotic fluid. *Bulletin of Environmental Contamination and Toxicology* 72:1226–31. PMID: 15362453.

Skakkebaek, N. E. 2002. Endocrine disrupters and testicular dysgenesis syndrome. *Hormone Research* 57 (Suppl. 2):43. doi:10.1159/000058100.

Stephens, M. L., K. Betts, N. B. Beck, V. Cogliano, K. Dickersin, S. Fitzpatrick, J. Freeman, G. Gray, T. Hartung, J. McPartland, et al. 2016. The emergence of systematic review in toxicology. *Toxicological Sciences* 152:10–16. doi:10.1093/toxsci/kfw059.

Suzuki, Y., J. Yoshinaga, Y. Mizumoto, S. Serizawa, and H. Shiraishi. 2012. Foetal exposure to phthalate esters and anogenital distance in male newborns. *International Journal of Andrology* 35:236–44. doi:10.1111/j.1365-2605.2011.01190.x.

Swan, S. H. 2008. Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. *Environmental Research* 108:177–84. PMID: 18949837.

Swan, S. H., K. M. Main, F. Liu, S. L. Stewart, R. L. Kruse, A. M. Calafat, C. S. Mao, J. B. Redmon, C. L. Ternand, S. Sullivan, et al. 2005. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environmental Health Perspectives* 113: 1056–61. PMID: 16079079.

Swan, S. H., S. Sathyanarayana, E. S. Barrett, S. Janssen, F. Liu, and R. H. Nguyen. 2015. First trimester phthalate exposure and anogenital distance in newborns. *Human Reproduction* 30:963–72. doi:10.1093/humrep/deu363.

Thankamony, A., N. Lek, D. Carroll, M. Williams, D. B. Dunger, C. L. Acerini, et al. 2014. Anogenital distance and penile length in infants with hypospadias or cryptorchidism: Comparison with normative data. *Environmental Health Perspectives* 122:207–11. doi:10.1289/ehp.1307178.

Viechtbauer, W. 2005. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* 30:261–93. doi:10.3102/10769986030003261.

Viechtbauer, W. 2010. Conducting meta-analysis in R with the metaphor package. *Journal of Statistical Software* 36:1–48. doi:10.18637/jss.v036.i03.

Vo, T. T., E. M. Jung, V. H. Dang, K. Jung, J. Baek, K. C. Choi, et al. 2009. Differential effects of flutamide and di-(2-ethylhexyl) phthalate on male reproductive organs in a rat model. *Journal of Reproduction and Development* 55: 400–11. PMID: 19367084.

Welsh, M., P. T. Saunders, M. Fisken, H. M. Scott, G. R. Hutchison, L. B. Smith, and R. M. Sharpe. 2008. Identification in rats of a programming window for reproductive tract masculinization, disruption of which leads to hypospadias and cryptorchidism. *Journal of Clinical Investigation* 118:1479–90. doi:10.1172/JCI34241.

Wilson, V. S., K. L. Howdeshell, C. S. Lambright, J. Furr, and L. E. Gray Jr. 2007. Differential expression of the phthalate syndrome in male Sprague-Dawley and Wistar rats after in utero DEHP exposure. *Toxicology Letters* 170:177–84. doi:10.1016/j.toxlet.2007.03.004.

Wohlfahrt-Veje, C., K. M. Main, and N. E. Skakkebaek. 2009. Testicular dysgenesis syndrome: Foetal origin of adult reproductive problems. *Clinical Endocrinology* 71:459–65. doi:10.1111/j.1365-2265.2009.03545.x.

Wolfe, G. W., and K. A. Layton. 2005. *Diethylhexylphthalate: Multigenerational reproductive assessment by continuous breeding when administered to sprague-dawley rats in the diet*. Gaithersburg, MD: TherImmune Research Corporation. TRC Study No. 7244-200. NTPRACB 98-004.

Zhang, L. D., Q. Deng, Z. M. Wang, M. Gao, L. Wang, T. Chong, and H. C. Li. 2013. Disruption of reproductive development in male rat offspring following gestational and lactational exposure to di-(2-ethylhexyl) phthalate and genistein. *Biological Research* 46:139–46. doi:10.4067/S0716-97602013000200004.