Male reproductive toxicity in epidemiology studies of phthalates: a case study application of systematic review approaches

Elizabeth Radke, Glinda Cooper

U.S. EPA, Office of Research and Development (ORD), National Center for Environmental Assessment (NCEA), IRIS Division

Introduction

Phthalates have anti-androgenic activity in rodents resulting in reduced circulating testosterone and male reproductive tract abnormalities. Several epidemiologic studies have examined this association in humans. The National Academies of Sciences (NAS) recently published a systematic review of endocrine-related low-dose toxicity that included examination of phthalates and male reproductive tract development, and the Integrated Risk Information System (IRIS) performed a systematic review of all male reproductive effects of phthalate exposure, following recommendations in the 2014 NAS review of the IRIS program. Here, we use the associations between anogenital distance (AGD) in humans and two phthalates, di(2-ethylhexl phthalate (DEHP) and diisobutyl phthalate (DIBP), as a case study of the IRIS systematic review process. We also compare our conclusions to those of the NAS and summarize our overall findings on epidemiology studies of male reproductive effects of phthalates.

Methods

Epidemiology studies were identified by conducting a single broad literature search on the six phthalates of interest. The following databases were searched: PubMed, Web of Science, and Toxline. The last update was in January 2017. Title/abstract and full text screening was performed by two reviewers. Studies were evaluated by at least two reviewers using the approach in Figure 1.

Domain judgments

judgment	Interpretation
Good	Appropriate study conduct relating to the domain & minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Poor	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that is interpreted to be the primary driver of any observed effect or makes the study uninterpretable. Study is not used without exceptional justification.

Individual study level domains

Epidemiology
Exposure measurement
Outcome ascertainment
Population Selection
Confounding
Analysis
Sensitivity
Selective reporting

Overall study rating

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity would be unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification

Figure 1. Study evaluation process

After study evaluation, the evidence for each outcome was synthesized for each phthalate, considering aspects of an association that may suggest causation. Based on this, the evidence was assigned within stream confidence judgments of *robust, moderate, slight, indeterminate,* or *compelling evidence of no effect.* The judgments for individual outcomes were summarized into an overall conclusion for male reproductive effects using a structured framework (see Poster by Yost et al.).

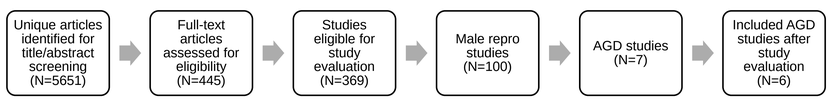

Figure 2. Abbreviated literature flow diagram

Results

Table 1. Epidemiology studies of AGD and phthalate exposure

	Reference	Study description			Study evaluation					
		Population	Exposure	Outcome	Exposure	Outcome	Selection	Confounding	Analysis	Overall confidence
Included	Bornehag et al., 2015	Birth cohort (N=196 boys) in Sweden	Single urine sample (1st trimester)	AGD at 19-21 mo	A/P	G	G	A	G	Medium
	Bustamante-Montes et al., 2013	Birth cohort (N=73 boys) in Mexico	Single urine sample (3rd trimester)	AGD at 1-2 d	P	G	A	A	G	Low
	Jensen et al., 2016	Birth cohort (N=273 boys) in Denmark	Single urine sample (26-30 wk gestation)	AGD at 3 mo	A/P	G	G	A	G	Medium
	Suzuki et al., 2012	Birth cohort (N=73 boys) in Japan	Single urine sample (3rd trimester)	AGD at 1-3 d	P	A	P	P	A	Low
	Swan, 2008	Birth cohort (N=106 boys) in U.S.	Single urine sample (mean 28 wk gestation)	AGD at 1-36 mo	A/P	P	G	A	P	Low
	Swan et al., 2015	Birth cohort (N=365 boys) in U.S.	Single urine sample(1st trimester)	AGD at 1-2 d	A/P	G	A	A	A	Medium

G=good; A=adequate; P=poor; A/P=adequate for short chain phthalates, poor for long chain.
Studies with biomarker measures based on samples other than urine (e.g., blood) were considered to be critically deficient for all short-chain phthalates and for primary metabolites (e.g., MEHP, MINP) of long-chain phthalates.

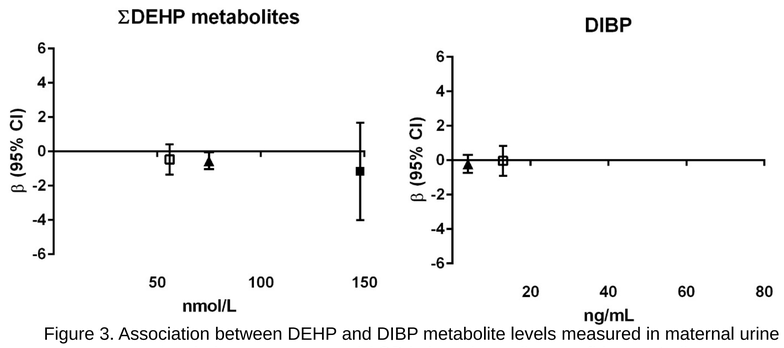

Figure 3. Association between DEHP and DIBP metabolite levels measured in maternal urine samples during pregnancy and AGD in boys in medium confidence studies

Regression coefficients on the y-axis are plotted against exposure level on the x-axis (population median for each study).

Table 2. Evidence profile table for epidemiology studies of AGD and DEHP and DIBP

	Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings and within stream evidence judgment
DEHP	**Medium confidence** Bornehag et al., 2015 Jensen et al., 2016 Swan et al., 2015 **Low confidence** Bustamante-Montes et al., 2013 Suzuki et al., 2012 Swan, 2008	Among medium confidence studies: • consistency • exposure-response gradient across studies • minimal concerns for bias	• low precision in study with largest effect size	⊕⊕◯ MODERATE Inverse associations between DEHP exposure and anogenital distance reported in 5/6 studies (Jensen et al., 2016, Swan et al., 2015, Bornehag et al., 2015, Swan, 2008, Suzuki et al., 2012), of which 2 were statistically significant (Swan et al., 2015, Swan, 2008). Among the 3 medium confidence studies, effect size increased with increasing exposure levels.
DIBP	**Medium confidence** Jensen et al., 2016 Swan et al., 2015 **Low confidence** Swan, 2008	• low study sensitivity may explain lack of association	• inconsistency • few studies	⊕◯◯ SLIGHT Inverse associations between DIBP exposure and anogenital distance reported in 2/3 studies (Swan, 2008, Swan et al., 2015), though neither were statistically significant. Exposure levels and range were low in all studies.

Of the seven identified studies on phthalates and AGD (Figure 2), one was excluded due to inadequate exposure measurement. Summary of the evaluations for the six included studies is in Table 1. Results of medium confidence studies were given priority (Figure 3), but all studies were included in the synthesis, which is summarized in the evidence profile table (Table 2). For DEHP, an exposure response gradient was observed across studies, with the study with the highest exposure levels reporting the strongest association. This was not observed for DIBP, but exposure levels were low in all studies. The same methods were used for other phthalate/outcome combinations and the within stream evidence judgments are shown in Figure 4. Table 3 presents a comparison of the within stream judgments from the IRIS and NAS reviews of anogenital distance, testosterone in infants, and hypospadias. Both found that the evidence for the latter two outcomes was not adequate to form a conclusion. For anogenital distance, evidence for DEHP and DBP was considered *moderate* in both reviews. Evidence for DINP, DIBP, and BBP was considered *slight* by IRIS and *inadequate* by NAS. These conclusions were not considered inconsistent, but rather reflect differences in the process for evidence synthesis. Only DEP differed between reviews, classified as *slight* by IRIS and *moderate* by NAS based on the results of a meta-analysis.

Outcome	DEHP	DINP	DBP	DIBP	BBP	DEP
Anogenital distance	M	S	M	S	S	S
Hypospadias/cryptorchidism	I	S	S	S	S	I
Pubertal development	S	S	S	S	S	S
Semen parameters	M	M	R	S	M	S
Time to pregnancy	S	I	M	S	M	I
Testosterone	M	M	S	M	I	I
Male repro overall	R	M	R	M	M	S
	Robust (R)	Moderate (M)	Slight (S)	Indeterminate (I)		

Figure 4. Within stream evidence judgments for human evidence of male reproductive effects associated with phthalates

Table 3. Within stream evidence judgments of systematic reviews of male reproductive developmental toxicity in epidemiology studies by IRIS and NAS

	Anogenital distance		Testosterone in infants		Hypospadias	
Phthalate	IRIS	NAS	IRIS	NAS	IRIS	NAS
DEHP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DINP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DBP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DIBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
BBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DEP	Slight	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate

Classifying levels: IRIS: Robust, Moderate, Slight, or Indeterminate; NAS: High, Moderate, Low, or Inadequate

Discussion

Overall, the results from epidemiology studies of male reproductive effects provide evidence of a hazard from phthalate exposure. Looking specifically at anogenital distance, there is *moderate* evidence of an association with DEHP and DBP exposure, and *slight* evidence for other phthalates. These findings are generally consistent with the NAS report on low-dose toxicity from endocrine active chemicals (2017). In the case of DIBP, the weaker evidence may be largely explained by the smaller number of studies and low exposure levels that decreased study sensitivity.