Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals

Committee on Endocrine-Related Low-Dose Toxicity

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Consensus Study Report of

The National Academies of

SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS

Washington, DC

www.nap.edu

THE NATIONAL ACADEMIES PRESS **500 Fifth Street, NW** **Washington, DC 20001**

This project was supported by Contract EP-C-14-005, Task Order #5, between the National Academy of Sciences and the US Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the view of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-45862-7
International Standard Book Number-10: 0-309-45862-5
Digital Object Identifier: https://doi.org/10.17226/24758

Additional copies of this publication are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: 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. doi: https://doi.org/10.17226/24758.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE ON ENDOCRINE-RELATED LOW-DOSE TOXICITY

Members

DAVID C. DORMAN (*Chair*), North Carolina State University, Raleigh, NC
WEIHSUEH CHIU, Texas A&M University, College Station, TX
BARBARA F. HALES, McGill University, Montreal, Quebec, Canada
RUSS B. HAUSER, Harvard T.H. Chan School of Public Health, Boston, MA
KAMIN J. JOHNSON, The Dow Chemical Company, Midland, MI
KAREN A. ROBINSON, Johns Hopkins University, Baltimore, MD
ANDREW A. ROONEY, National Toxicology Program, Research Triangle Park, NC
RUTHANN RUDEL, Silent Spring Institute, Newton, MA
SHEELA SATHYANARAYANA, University of Washington, Seattle, WA
SUSAN L. SCHANTZ, University of Illinois at Urbana-Champaign, IL
KATRINA WATERS, Pacific Northwest National Laboratory, Richland, WA

Staff

SUSAN N. J. MARTEL, Project Director
ELLEN K. MANTUS, Scholar and Director of Risk Assessment
LINDA O'DOUGHDA, Editor
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects
TAMARA DAWSON, Program Coordinator

Sponsor

US ENVIRONMENTAL PROTECTION AGENCY

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Members

WILLIAM H. FARLAND (*Chair*), Colorado State University, Fort Collins, CO
RICHARD A. BECKER, American Chemistry Council, Washington, DC
E. WILLIAM COLGLAZIER, American Association for the Advancement of Science, Washington, DC
DOMINIC M. DITORO, University of Delaware, Newark, DE
DAVID C. DORMAN, North Carolina State University, Raleigh, NC
CHARLES T. DRISCOLL, JR., Syracuse University, Syracuse, NY
ANNE FAIRBROTHER, Exponent Inc., Philomath, OR
GEORGE GRAY, The George Washington University, Washington, DC
STEVEN P. HAMBURG, Environmental Defense Fund, New York, NY
ROBERT A. HIATT, University of California, San Francisco, CA
SAMEUL KACEW, University of Ottawa, Ontario
H. SCOTT MATTHEWS, Carnegie Mellon University, Pittsburgh, PA
ROBERT PERCIASEPE, Center for Climate and Energy Solutions, Arlington, VA
R. CRAIG POSTLEWAITE, Burke, VA
MARK A. RATNER, Northwestern University, Evanston, IL
JOAN B. ROSE, Michigan State University, East Lansing, MI
GINA M. SOLOMON, California Environmental Protection Agency, Sacramento, CA
ROBERT M. SUSSMAN, Sussman and Associates, Washington, DC
DEBORAH L. SWACKHAMMER, University of Minnesota, St. Paul, MN
PETER S. THORNE, University of Iowa, Iowa City, IA

Senior Staff

TERESA A. FRYBERGER, Director
ELLEN K. MANTUS, Scholar and Director of Risk Assessment
RAYMOND A. WASSEL, Scholar and Director of Environmental Studies
SUSAN N. J. MARTEL, Senior Program Officer for Toxicology

Acknowledgments

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report:

Sir Colin Berry, University of London
Kim Boekelheide, Brown University
Steven P. Bradbury, Iowa State University
Jonathan Chevrier, McGill University
Deborah Cory-Slechta, University of Rochester
George P. Daston, Proctor & Gamble Company
Brenda Eskenazi, University of California, Berkeley
Dale Hattis, Arlington, MA
Malcolm Macleod, University of Edinburgh
Bruce McEwen, The Rockefeller University
David Savitz, Brown University
Laura Vandenberg, University of Massachusetts, Amherst
Tracey Woodruff, University of California, San Francisco
Yiliang Zhu, University of South Florida

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report nor did they see the final draft before its release. The review of this report was overseen by Gary Ginsberg, Connecticut Department of Public Health, and Martin Philbert, University of Michigan. They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

The four protocols for conducting the systematic reviews contained in this report were also peer reviewed before the systematic reviews were undertaken. We thank the following individuals for their review of the protocols:

Sir Colin Berry, University of London
Kim Boekelheide, Brown University
Deborah Cory-Slechta, University of Rochester
Brenda Eskenazi, University of California, Berkeley
Malcolm Macleod, University of Edinburgh
John Meeker, University of Michigan
Tracey Woodruff, University of California, San Francisco

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the protocols, nor did they see the final protocols. The review of the protocols was overseen by David Eaton, University of Washington, and Martin Philbert, University of Michigan.

The committee is grateful to Juleen Lam and Tracey Woodruff at the University of California, San Francisco, for sharing a draft of their systematic review on PBDE (polybrominated diphenyl ether) exposure and neurodevelopmental effects in childhood. Their generosity and cooperation allowed the committee to demonstrate the methods by which a systematic review can be reviewed and updated. It is a fortunate development that our report has been published almost simultaneously with theirs.

The committee gratefully acknowledges the following individuals for their presentations to the committee during open sessions or their participation in a workshop: Stanley Barone, US Environmental Protection Agency; Kim Boekelheide, Brown University; Joseph Braun, Brown University; Germaine Buck-Louis, National Institute of Child Health and Human Development; Thomas Burke, US Environmental Protection Agency; Vincent Cogliano, US Environmental Protection Agency; Glinda Cooper, US Environmental Protection Agency; David Dix, US Environmental Protection Agency; Daniel Doerge, US Food and Drug Administration; Brenda Eskenazi, University of California, Berkeley; Jodi Flaws, University of Illinois at Urbana-Champaign; Catherine Gibbons, US Environmental Protection Agency; Earl Gray, US Environmental Protection Agency; and John Meeker, University of Michigan. The committee also thanks Robyn Blain and Pamela Hartman at ICF International for their assistance with data extraction, 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, and Anne Johnson for her editorial assistance and advice.

Contents

APPENDIXES

BOXES, FIGURES, AND TABLES

BOXES

FIGURES

TABLES

Abbreviations and Acronyms

ADHD	attention deficit/hyperactivity disorder
ADR	adverse drug reaction
AGD	anogenital distance
AGD ap	anogenital distance (anopenile)
AGD as	anogenital (anoscrotal)
AGI	anogenital index
AHRQ	Agency for Healthcare Research and Quality
AICc	Akaieke information criterion corrected
AOP	adverse outcome pathway
APD	anopenile distance
ASD	anoscrotoal distance
BASC-2	Behavioral Assessment System for Children-2
BDE	brominated diphenyl ether
BMD	benchmark dose
BMDL	benchmark dose modeling
BMR	benchmark response
BPA	bisphenol A
BzBP	benzylbutyl phthalate
CADS-P	Conners' ADHD Index (derived from Conners' Parent Rating Scale)
CASTNET	Clean Air Status and Trends Network
CBCL	Child Behavior Checklist
CDC	US Centers for Disease Control and Prevention
CI	confidence interval
COI	conflict of interest
cpAOP	computationally predicted adverse outcome pathway
CPRS	Conners' Parent Rating Scale-Revised
CPT II	Conners' Continuous Performance Test II
CRS-T	Conners' Rating Scale-Teachers
DBD	Disruptive Behavior Disorders Rating Scale
DBP	dibutyl phthalate
DEHP	di(2-ethylhexyl) phthalate
DEP	diethyl phthalate
DES	diethylstilbestrol
DIBP	diisobutyl phthalate
DIDP	diisodecyl phthalate
DINP	diisononyl phthalate
DIOP	diisooctyl phthalate
DMP	dimethyl phthalate
DNOP	di-n-octylphthalate
DPP	dipentyl phthalate
EAC	endocrine active chemical
EDSP	Endocrine Disruptor Screening Program
EFSA	European Food Safety Authority

EPA	US Environmental Protection Agency
ExpoCast	EPA's computer model to predict chemical exposure
FAERS	FDA Adverse Event Reporting System
FDA	US Food and Drug Administration
FQPA	Food Quality Protection Act (of 1996)
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HAWC	Health Assessment Workspace Collaborative
INSL-3	insulin-like 3
IRIS	Integrated Risk Information System
K-CPT	Conners' Kiddie Continuous Performance Test
LOAEL	lowest-observed-adverse-effect level
LOEL	lowest-observed-effect level
MBP	monobutylphthalate
MBzP	mono-benzyl phthalate
MCNP	mono-(carboxynonyl) phthalate
MCOP	mono-carboxy-isooctyl phthalate
MCPP	mono-(3-carboxypropyl) phthalate
MECPP	mono-(2-ethyl-5-carboxypentyl) phthalate
MEHP	mono-2-ethylhexyl phthalate
MEP	mono-ethyl phthalate
MeO-BDE	methoxylated BDE
MeSH	medical subject heading
MIBP	mono-isobutyl phthalate
MMP	mono-methyl phthalate
MSCA	McCarthy Scales of Children's Abilities
NHANES	National Health and Nutrition Examination Survey
NMDR	nonmonotonic dose-response
NOAEL	no-observed-adverse-effect level
NOEL	no-observed-effect level
NRC	National Research Council
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OH-BDE	hydroxylated BDE
OHAT	National Toxicology Program's Office of Health Assessment and Translation
PBDE	polybrominated diphenyl ether
PBPK	physiologically based pharmacokinetic
PCB	polychlorinated biphenyl
PECO	Population, Exposure, Comparator, and Outcome
PM	particulate matter
PND	postnatal day
QSAR	quantitative structure-activity relationship
RE	random effects
RfD	reference dose
ROM	ratio of means
RR	relative risk
SD	standard deviation
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
ToxCast	Toxicity Forecaster
VAERS	Vaccine Adverse Event Reporting System
WHO	World Health Organization
WISC	Wechsler Intelligence Scale for Children
WPPSI	Wechsler Preschool and Primary Scale of Intelligence

Summary

To safeguard public health, the US Environmental Protection Agency (EPA) must keep abreast of new scientific information and emerging technologies so that it can apply them to regulatory decision making. For decades the agency has dealt with questions about what animal-testing data to use to make predictions about human health hazards, how to perform dose-response extrapolations, how to identify and protect susceptible subpopulations, and how to address uncertainties. As alternatives to traditional toxicity testing have emerged, the agency has been faced with additional questions about how to incorporate data from such tests into its chemical assessments and whether such tests can replace some traditional testing methods. In addition, evidence has emerged suggesting that some chemicals have effects at doses lower than those used in traditional toxicity testing, raising concerns that traditional toxicity-testing protocols might be inadequate to identify all potential hazards to human health. In particular, endocrine active chemicals (EACs), or endocrine disruptors, have been a focal point for these questions because they have the ability to modulate normal hormone function, and small alterations in hormone concentrations, particularly during sensitive life stages, can have lasting and significant effects.

To address concerns about potential human health effects from EACs at low doses, EPA requested that the National Academies of Sciences, Engineering, and Medicine develop a strategy to evaluate the evidence for such low-dose effects (see Box S-1). The National Academies convened an ad hoc committee of experts to address this task. The task specified that the committee should perform systematic reviews of animal and human studies on at least two chemicals and demonstrate how the results can be integrated and considered with other relevant data to draw conclusions about causal associations. This report describes the strategy developed by the committee and highlights the role systematic review methods play in this overall strategy. The result of the systematic reviews and the lessons learned in performing them are also presented in the report.

BOX S-1 Statement of Task

An ad hoc committee under the auspices of the National Research Council (NRC) will develop a strategy for evaluating evidence of low-dose adverse human effects that act through an endocrine-mediated pathway. The study will include a scientific workshop to support the conduct of systematic reviews of human and animal toxicology data for two or more chemicals that affect the estrogen, androgen, or perhaps other endocrine systems. The workshop will seek to identify examples of relevant chemicals, populations/model systems, and end points of interest for further study using systematic review methods. Systematic reviews for these chemicals/populations/end points for human and animal data streams will be performed under the direction of the committee. The committee will evaluate the results of the systematic reviews, demonstrate how human and animal data streams can be integrated, determine whether the evidence supports a likely causal association, and evaluate the nature and relevance of the dose-response relationship(s). The committee will consider how to use adverse outcome pathway (AOP) or other mechanistic data, including high-throughput data and pharmacokinetic information, to elucidate under what circumstances human and animal data may be concordant or discordant. The results of the committee's evaluation of low-dose toxicity can be used to inform EPA on the adequacy of its current regulatory toxicity-testing practices.

STRATEGY FOR EVALUATING LOW-DOSE EFFECTS

The committee developed a generic strategy for evaluating evidence of low-dose[1] effects that includes three broad phases: surveillance for signals that a chemical may cause a health effect or that a health effect may be missed by traditional toxicity-testing methods, investigation and analysis of the evidence, and acting on the evidence (see Figure S-1). The first two phases involve identifying issues or questions to address, determining the best methods for evaluating them, and then conducting appropriate investigation and analyses to support the type of decision to be made. In its deliberations, the committee considered and demonstrated how these two phases apply to the evaluation of EACs. The last phase of the strategy involves policy and other management decisions that fall outside of the committee's task.

Surveillance

In the strategy, surveillance refers to a process for detecting signals that raise questions about the potential low-dose toxicity of a particular chemical or about the ability to detect low-dose toxicity more generally. For example, signals might include an indication that an adverse outcome in a human population could be related to an EAC exposure, or evidence that a particular low-dose effect might not be detectable with traditional toxicity testing. To conduct the surveillance necessary to identify such signals, the committee identified three broad categories of data that should be monitored on a regular basis. These include data on specific chemicals, information that could have implications for toxicity-testing methods and best practices for EACs, and information on endocrine-related disease in animals and humans. Such information could be obtained by conducting regular surveys of the scientific literature, gathering input from stakeholders, and collecting information about human exposure, for example, through biomonitoring data, external exposure measurements, and computational models that link external and internal exposure.

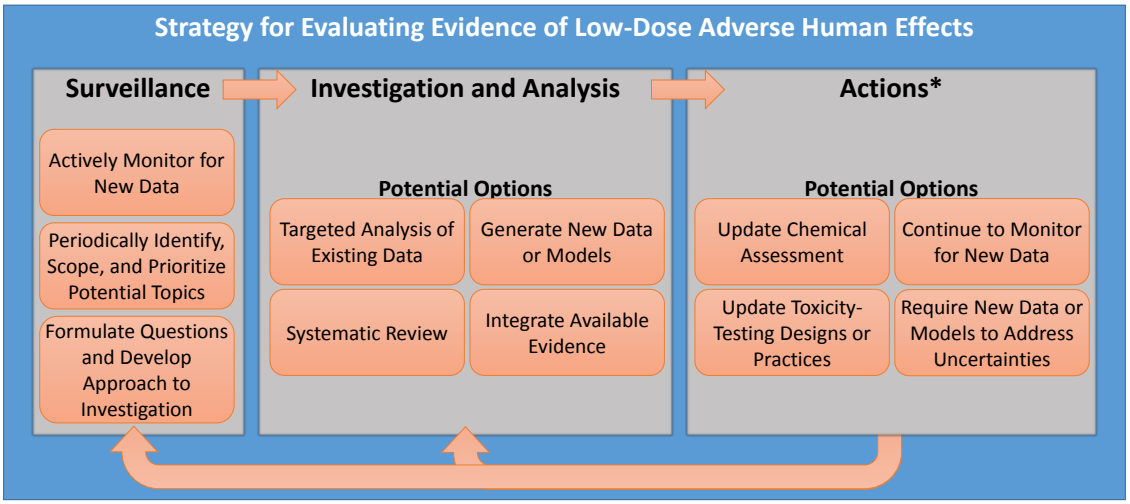


FIGURE S-1 Strategy for evaluating evidence of adverse human effects from low-dose exposure to chemicals. The strategy includes three broad phases: surveillance, investigation and analysis, and actions. Each phase includes multiple options that may be employed alone or in combination. The order in which the options appear does not indicate a hierarchy or a sequence that should be followed. *Recommendations for this phase of the strategy were outside of the committee's charge.

[1]Low dose is defined in the report as external or internal exposure that falls within the range estimated to occur in humans. Human exposure estimates may be based on environmental or biomarker measurements and/or computational models. If no human exposure data are available, low dose is defined on a case-by-case basis relative to an explicitly defined exposure in a particular context.

Once signals are identified, scoping exercises can be used to prioritize areas for investigation and analysis. Scoping involves using the scientific literature and other information to determine the extent, range, and nature of the information on the topic, to identify data gaps, and to consider what additional analyses are needed. Additional factors that could influence the decision to pursue a topic are the size of the population at risk, the public health significance of the issue, and available resources. The scoping exercise also considers the potential actions that might be taken in response to the signal, which will help determine the level of scientific depth and rigor that might be required to inform any such actions.

When a signal is prioritized for further investigation, the next step is to formulate key questions to frame the issues. Once key questions are identified, it is possible to design an approach to answer those questions using appropriate tools for investigation and analysis.

Investigation and Analysis

The committee outlines four main options for investigation and analysis aimed at understanding the potential human health effects from exposure to EACs at low doses: targeted analysis of existing data, generation of new data or models, systematic review of evidence, and integration of evidence. The approaches used should be selected on a case-by-case basis; in some cases one approach will be sufficient while in other cases several investigative approaches might be needed to adequately answer the questions.

Targeted Analysis of Existing Data

Targeted analysis is a method for analyzing (or reanalyzing) data. It can allow for better comparison of results between studies. For example, when outcomes are measured differently between studies (e.g., as continuous or dichotomous variables), it might be possible to convert the data to allow for comparisons. Statistical and other computational approaches to characterize dose-response relationships may be used to provide evidence of low-dose effects. Qualitative analyses can also be used to make judgments about seemingly discordant data. For example, if effects are seen at different doses in two or more studies, it is useful to evaluate the studies for factors that could explain the differences.

Generation of New Data or Models

Some questions can be addressed only through generation of new data or the development of new methods to analyze data. In cases where new data are needed, the investigation and analysis phase would focus on determining the type of data needed to characterize the human health effects and the best methods for obtaining the data. These efforts could include the experimental studies to fill data gaps or the development of new computational models (e.g., physiologically based pharmacokinetic models to address questions about dosimetry or species differences).

Systematic Review

For cases where a rigorous assessment is needed to address a question, a systematic review can help focus the evaluation and maximize transparency in both how the assessment was conducted as well as how it was used to draw conclusions. To help ensure that the evidence is selected and evaluated in an objective and consistent manner, a systematic review requires carefully crafting the research question and planning in advance what methods will be used for screening and analyzing the scientific literature to answer the question.

Integration of Evidence

The fourth type of analysis is integration of available evidence to draw conclusions. Evidence integration largely focuses on hazard identification. Drawing conclusions about hazards related to low-dose

effects of EACs will typically also require evaluating evidence specifically on the nature of the dose-response relationship at low doses. Environmental exposure data, such as biomonitoring data, can be useful in defining what subset of data can be considered relevant to low-dose exposures. Evidence integration to address questions about low-dose effects might also need to consider in vitro and mechanistic evidence, modeled dose-response relationships, and co-exposures.

Data integration can also be used to address broader questions such as whether a "new" end point or "new" exposure or assessment window is relevant to determining low-dose toxicity. For example, some end points have been added to regulatory testing protocols in response to growing evidence that they are indicators of toxicity, and the duration of some tests has been extended to capture effects that might occur later in life. Signals identified during the surveillance phase that have these types of implications about toxicity testing should be evaluated by integrating the available evidence.

Actions

Once the investigation and analysis phase has been completed, the next step is to select the type of action (or actions) warranted. As shown in Figure S-1, several options for action could be appropriate, including updating chemical assessments, continuing to monitor for new data, updating toxicity-testing designs and practices, or requiring new data or models to reduce uncertainties. The type of action that EPA takes could be influenced by additional factors, including existing policies and regulations, the size of the population at risk, the public health significance of the human health effects, and available resources. Making recommendations about what actions to take was outside the scope of this committee's activities.

Findings and Recommendations

To ensure adequate understanding of hazards and to inform its decisions about its regulatory toxicity-testing practices, EPA needs a general strategy for ongoing evaluation of evidence of low-dose effects from exposure to EACs. The committee proposes a strategy involving three phases: surveillance, investigation and analysis, and actions. EPA is already conducting many activities consistent with the proposed strategy, though not necessarily in the specific context of assessing low-dose exposure to EACs.

Recommendation: EPA should develop an active surveillance program focused specifically on low-dose exposures to EACs. This program could include regularly monitoring published research and other information sources, gathering input from stakeholders, and considering human exposure information. It might also involve data collection in collaboration with other agencies and outside parties. The surveillance program should periodically identify, scope, and prioritize potential areas of focus related to low-dose effects, such as particular chemicals and end points. Some approaches will require methods and tool development, such as automated methods for monitoring the literature.

Recommendation: After a topic is selected for further evaluation, the agency should plan its investigation by identifying key questions to be addressed and determining the types of data and analyses needed to answer the questions and to support future agency actions. The specific approaches and tools used to implement the strategy to address issues related to low-dose endocrine effects will need to be considered on a case-by-case basis and should be guided by the questions under study.

The four main options for investigation and analysis include targeted analysis of existing data, generation of new data or models, systematic review, and integration of evidence. Different approaches will be appropriate for different circumstances. The types of analyses used to investigate the questions are not mutually exclusive, and in some cases several approaches might be needed to address the questions adequately. Integration of evidence for low-dose adverse human effects of EACs involves consideration of both hazard identification and dose response.

Recommendation: Human environmental exposure or biomonitoring data should be used, if available, to define what subset of the data should be considered as reflective of low-dose exposure.

The proposed strategy will facilitate a greater emphasis on regular consideration of the adequacy of toxicity testing for assessing low-dose exposures to EACs. However, the agency will also be faced with questions about the amount and quality of evidence needed in order to justify updating test methods, and these questions might be more appropriately addressed through policy decisions.

IMPLEMENTING THE STRATEGY: EXAMPLE REVIEWS

In its charge to the committee, EPA requested that the committee perform systematic reviews of animal and human studies on at least two chemicals and show how the results from the animal and human evidence can be integrated to draw conclusions. Systematic reviews and integration of evidence are two of the four options for further investigating and analyzing topics of interest in phase two of the committee's proposed strategy. The committee undertook these example reviews to demonstrate how these approaches could be used in a strategy to evaluate low-dose toxicity of EACs and to identify lessons learned that could help EPA employ these approaches successfully. However, systematic reviews and integration of evidence will not be appropriate or required in all circumstances.

To select EACs for its example reviews, the committee conducted several exercises to illustrate how phase one of the strategy (Surveillance) might be performed for a question ("Is there evidence of low-dose adverse human effects that act through an endocrine-mediated pathway?"). These surveillance exercises included garnering stakeholder input through a public workshop, surveying the scientific literature, and collecting information about human exposure. As part of phase one, the committee prioritized candidate chemicals using criteria aimed at addressing the elements set forth in the statement of task. For example, because the committee was tasked with demonstrating how different evidence streams can be integrated, it purposely selected EACs for which there appeared to be an adequate number of animal and human studies to allow for comparisons and integration. The two EACs chosen were phthalates and polybrominated diphenyl ethers (PBDEs). Before undertaking its reviews, the committee refined key questions about the effects of the selected chemicals.

At the start of phase two (Investigation and Analysis), the committee developed protocols to use animal and human studies to answer those questions posed in phase one. Protocols were based on the methods developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) and were peer reviewed before the systematic reviews were undertaken. The protocols identified methods of analysis (e.g., meta-analysis) the committee would use. The systematic review method included a framework for drawing conclusions about the "level of evidence" for a health effect as being inadequate, low, moderate, or high. Level-of-evidence ratings were subsequently used in the evidence integration step to classify the hazard associated with a given chemical as not classifiable, suspected, presumed, or known (see Figure S-2). Mechanistic evidence was also considered in determining the hazard conclusion.

The first set of systematic reviews focused on the question of how phthalates might affect male reproductive-tract development. Phthalates[2] are ubiquitous environmental contaminants, and human exposure to them has been well documented. Phthalates are known to affect the androgen hormone system, which plays a critical role in the development of the male reproductive tract. The committee focused its investigation on three end points considered to be indicative of changes in androgen levels—anogenital distance (AGD), fetal testosterone levels, and hypospadias. The committee conducted separate reviews of

[2]The phthalates include benzylbutyl phthalate (BzBP), dibutyl phthalate (DBP), diethylhexyl phthalate (DEHP), diethyl phthalate (DEP), diisobutyl phthalate (DIBP), diisodecyl phthalate (DIDP), diisononyl phthalate (DINP), diisooctyl phthalate (DIOP), dimethyl phthalate (DMP), di-n-octyl phthalate (DOP), and dipentyl phthalate (DPP).

the animal and human evidence on phthalates and then integrated the evidence to draw conclusions about potential hazards, low-dose effects, and the adequacy of toxicity-testing methods for evaluating those hazards.

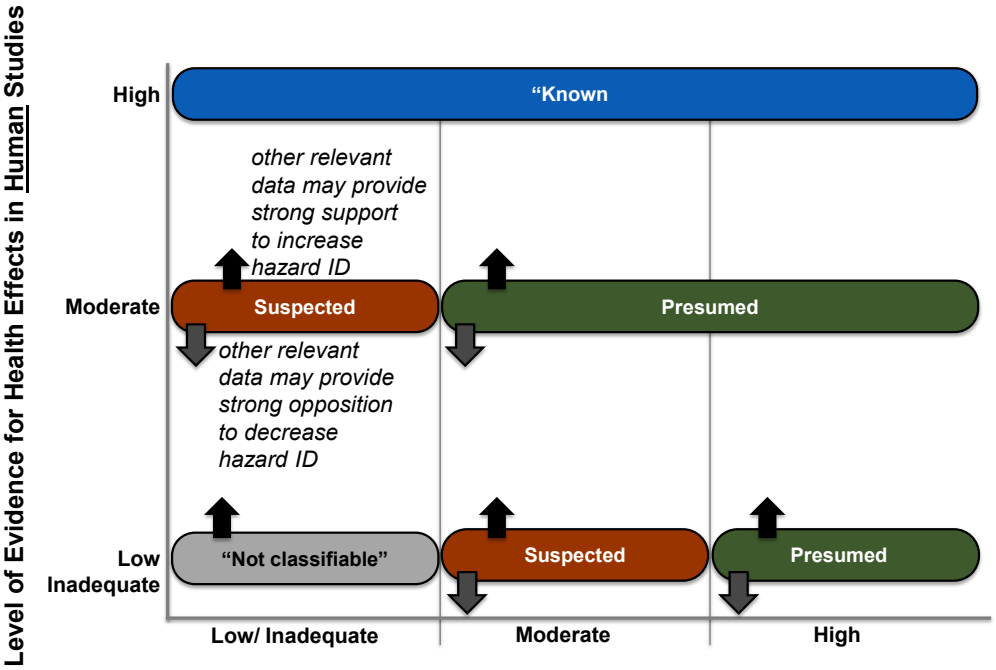


FIGURE S-2 OHAT hazard identification scheme. SOURCE: NTP (2015).

Effects of Phthalates on Male Reproductive-Tract Development

BOX S-2 Example 1

Question: What is the effect of in utero exposure to phthalates on male reproductive-tract development?

Example chemical examined: Diethylhexyl phthalate (DEHP)

Example end point examined: Anogenital distance (AGD), which is considered to be an indicator of reduced fetal androgen production.

Level-of-evidence conclusions: There is a high level of evidence from animal studies and a moderate level of evidence from human studies that fetal exposure to DEHP is associated with decreases in AGD.

Hazard conclusion: In utero exposure to DEHP is presumed to be a reproductive hazard to humans. This conclusion means that there was sufficient animal and human evidence to allow the committee to conclude that DEHP is a potential hazard to human health. Identifying the potential of a chemical to cause particular forms of toxicity in humans does not indicate whether the substance poses a risk in specific exposed populations. Such a determination requires the completion of a risk assessment that takes into consideration exposure of a given population; a risk assessment was not performed by the committee.

The landscape of data on phthalates is complex. There are varying amounts of data available for different phthalates, and a further complication is that human studies often involve exposure to mixtures of phthalates. Although the committee analyzed the evidence related to multiple phthalates and multiple end points, for the purposes of this summary the committee chose to highlight as an illustrative example its analysis of the association between diethylhexyl phthalate (DEHP) and AGD.[3]

The committee's systematic review of the relationship between DEHP and changes in AGD included 19 studies in animals and five studies in humans. Both animal and human studies showed reductions in male AGD after in utero exposure to DEHP. A meta-analysis of the animal studies found consistent evidence of a decrease in AGD in association with DEHP treatment, with a dose-response gradient. A meta-analysis of the human studies similarly found consistent evidence that increased maternal urinary concentrations of DEHP metabolites were associated with decreased AGD in male children. Overall, this evidence of an association between DEHP and decreased AGD supports the conclusion that in utero exposure to DEHP is presumed to be a reproductive hazard to humans.[4]

The committee considered whether pharmacokinetic or mechanistic data would influence this hazard conclusion. The committee found that mechanistic data from in vitro studies and animal models provide biological plausibility that exposure to DEHP is associated with a reduction in AGD in humans, based on decreased fetal testosterone as an intermediate effect. Moreover, androgen-dependent development of the male reproductive tract and the androgen dependence of AGD appear to be well conserved across mammalian species (including humans). Thus, the committee concluded that the pharmacokinetic and mechanistic data support the hazard conclusion that in utero exposure to DEHP is presumed to be associated with decreased AGD in humans.

Drawing conclusions about dose response is more challenging. It is difficult to directly compare the effects of different levels of DEHP exposure in animals and humans because animal studies typically report administered doses whereas studies in humans rely on the measurement of DEHP metabolites in urine or other body fluids. Some investigators have used pharmacokinetic models to estimate human daily intakes of DEHP based on concentrations of phthalate metabolites in urine; these models have suggested that human intake is markedly lower than the doses used in animal studies. However, differences in internal measures of exposure (concentrations in urine or amniotic fluid) were of a much smaller magnitude. Thus, the issue of phthalate effects on male reproductive-tract development represents an example where current toxicity-testing methods can identify a hazard that is presumed to be of concern to humans, but current methods might not be able to accurately predict exposures at which humans are affected. This finding also provides additional support for EPA's decision to include AGD measurements in regulatory toxicity testing.

The second set of systematic reviews focused on the question of how PBDEs might affect neurobehavioral function. PBDEs are ubiquitous in the environment, and human exposure to them has been well documented. The committee conducted its own review of available animal studies and updated a recent systematic review of human studies conducted by Lam et al., which was shared with the committee in draft form.[5] The review of the human studies evaluated effects on intelligence, attention deficit/hyperactivity disorder (ADHD), and attention-related behavioral conditions. For its review of animal studies, the committee focused on findings related to learning, memory, and attention which the committee considered to have the closest parallels to the intelligence and attention-related outcomes measured in the human studies. The review of animal studies included any type of PBDE, but the review of human studies was restricted to the types of PBDEs most commonly reported in human biological samples: BDE-47, -99, -100, and -153.

[3]A full analysis of other phthalates and end points is presented in Chapter 3. The hazard conclusions reached on the other phthalates and other end points were either equivalent to or weaker than the one reached for DEHP and AGD.

[4]Committee conclusions concerning DEHP effects on fetal testosterone and hypospadias rested on animal evidence since insufficient human evidence was available to assess whether exposure to DEHP is associated with these outcomes.

[5]The review was subsequently updated by the authors and is in press for publication in *Environmental Health Perspectives*.

Although the committee analyzed the evidence related to multiple brominated diphenyl ethers (BDEs), for the purposes of this summary the committee chose to focus on its analysis of the potential effects of BDE-47 exposure on learning and intelligence.[6]

The animal data on PBDEs and learning, memory, and attention were diverse and complex, with studies using varying designs, outcomes, and types of PBDEs. Six studies of BDE-47 and learning were available, and five found some indication of an effect on at least one measure of learning. The committee also conducted a meta-analysis by combining data from studies on PBDEs reporting latency in the Morris water maze, a test commonly used in studies of learning. The meta-analysis found consistent evidence of an increase in latency in the last trial of the Morris water maze in PBDE-exposed animals (meaning that the exposed animals took longer to locate the escape platform than nonexposed animals), and this effect was robust to multiple sensitivity analyses. The analysis also showed some evidence of a dose-response gradient, but these findings were inconsistent across studies. Differences among studies with regard to dose response might be due to variations in study design, such as the use of different PBDEs, differences in the duration of exposure, differences in internal dose, and differences in strain and species, or to other factors, such as potency of the congeners or pharmacokinetics.

To assess the human evidence, the committee critically evaluated the methods of a recent systematic review conducted by Lam et al. (in press) using an evaluation tool called ROBIS. Judging that this existing review fulfilled the requirements of a systematic review (it followed the Navigation Guide method for performing systematic reviews, which is similar to the OHAT method) and that there was no evidence of risk of bias in the assessment, the committee used the Lam et al. review as a basis for its own assessment. The authors of the review identified nine studies that measured IQ in relation to developmental PBDE exposure. In a meta-analysis of a subset of the studies, the authors found evidence of an association between PBDE exposure and a decrease in IQ. The committee conducted an updated literature search based on this review; finding no studies with substantively new findings, the committee determined that the conclusions of the Lam et al. systematic review would form a sufficient basis for the committee's work of integrating the available evidence.

Effects of PBDEs on Neurobehavioral Function

BOX S-3 Example 2

Question: Is developmental exposure to PBDEs associated with effects on neurobehavioral function?

Example chemical examined: BDE-47

Example end point examined: Learning in animals and intelligence in humans

Level of evidence conclusions: There is a moderate level of evidence that exposure to BDE-47 is associated with decrements in learning in rodents and decreases in IQ in humans.

Hazard classification conclusion: Overall, the evidence supports a conclusion that BDE-47 is a presumed hazard to humans with respect to effects on intelligence. This conclusion means that there was sufficient animal and human evidence to allow the committee to conclude that BDE-47 is a potential hazard to human health. Identifying the potential of a chemical to cause particular forms of toxicity in humans does not indicate whether the substance poses a risk in specific exposed populations. Such a determination requires the completion of a risk assessment that takes into consideration exposure of a given population; a risk assessment was not performed by the committee.

[6]A full analysis of other BDEs and end points is presented in Chapter 4. The hazard conclusions reached on the other congeners and end points were either equivalent to or weaker than the one reached for BDE-47 and learning in animals and intelligence in humans.

Reviewing mechanistic data on PBDEs in relation to developmental neurotoxicity, the committee identified some biological plausibility of the associations observed between PBDE exposure during the perinatal period and later neurobehavioral outcomes. However, the complexity and multifactorial nature of how PBDEs might affect neurodevelopmental processes hindered the committee's attempt to define an adverse outcome pathway.

As with phthalates, it is difficult to directly compare PBDE exposure in animal studies to that occurring in humans because the majority of animal studies report only administered doses whereas human studies rely on the measurement of PBDEs in serum or other body fluids. Estimates of human daily intakes based on measurements of PBDEs in food and dust suggest that human exposure is several orders of magnitude lower than that achieved with benchmark doses estimated from the data or the meta-analysis of the animal studies on PBDEs. Studies of internal doses of BDE-47 also show large disparities in the level of exposure between humans and animals, though these disparities are less pronounced than those suggested by the intake data. The available data on these measures were scant and uncertain, though, limiting the ability to use animal studies to predict exposure levels at which effects occur in humans. Thus, this is another situation in which current toxicity-testing methods can identify a hazard that is presumed to be of concern to humans, but current methods might not be able to accurately predict exposures at which humans are affected.

Findings and Recommendations

The following findings and recommendations stem from lessons learned in the committee's process of performing the systematic reviews and integrating evidence for the selected EACs. Additional findings and recommendations that are more specific to the selected EACs are provided in Chapters 3 and 4.

Findings Related to Conducting Systematic Reviews

Consistency and Transparency: The committee found that the systematic review process was valuable because it provided a framework for identifying, selecting, and evaluating evidence in a consistent and explicit manner; maximized transparency in how the assessments were performed; and facilitated the clear presentation of the basis for scientific judgments.

Chemical Mixtures: The two examples that the committee selected for its systematic reviews involved chemical classes rather than individual chemicals. In retrospect, this aspect added complexity to the reviews. In its evaluation of the phthalates, the committee evaluated individual phthalates separately, demonstrating how systematic reviews can be performed on single chemicals. In its evaluation of PBDEs, the committee considered different PBDEs both separately and in combination, demonstrating one way systematic reviews can be applied to chemical mixtures.

The Use of Meta-Analyses: The committee found that meta-analyses were valuable in summarizing data from the systematic reviews and in comparing the animal and human evidence in a robust and consistent manner. Meta-analyses can be used to inform confidence ratings for bodies of evidence and to support benchmark dose modeling.

> Recommendation: Systematic reviews should include meta-analysis of the animal and human evidence, if appropriate. The results of meta-analyses should be used to examine quantitative relationships between EACs and end points of interest to inform the confidence ratings of the bodies of evidence, and, if possible, to estimate benchmark doses.

Evaluating Risk of Bias: The committee found that information important to evaluating the quality of individual animal studies was often not reported, including whether the study controlled for litter effects, whether animals were randomly allocated to study groups, and whether research personnel were blinded

to the study groups during the outcome assessment. Because a lack of adequate reporting could not be distinguished from failure to adhere to practices that minimize bias, failure to report practices that minimize bias often led to higher risk of bias ratings for individual studies, downgrading the overall level of confidence in the body of evidence. These types of problems could be remedied if journals required better reporting of the methods used in animal studies, especially reporting pertaining to issues that might introduce bias into the research. These requirements could build on reporting standards that have been developed by various organizations to improve transparency (e.g., the ARRIVE guidelines). For example, studies should be required to report whether animals were assigned to study groups using random allocation and whether researchers were blinded to the study groups during outcome assessment.

The Use of an Existing Systematic Review: In the PBDE assessment, the committee found it was time saving to use a recent systematic review of the effect of developmental exposure to PBDEs on IQ and ADHD as a basis for its own assessment.

> Recommendation: EPA should develop policies and procedures to allow the agency to use and update existing systematic reviews. It is important that the existing systematic review's study question directly addresses EPA's topic of interest and that the methods are critically evaluated before the systematic review is used and updated.

Expertise Required: The committee found that conducting a systematic review and integrating evidence requires a multidisciplinary approach tailored to the specific review question. In particular, it is essential to have expertise in the conduct of meta-analyses and benchmark dose modeling.

Findings Related to Integrating Evidence

The committee found comparing evidence on dose-response relationships between animal and human studies to be challenging and imprecise because animal studies often report external administered doses (usually without measures of internal dose) whereas human studies measure biomarkers of internal dose (with estimates of the external administered dose being uncertain). Toxicology studies that measure internal dose metrics, including metrics that are similar to those used in human biomonitoring, could help address this data gap.

> Recommendation: To support animal-to-human extrapolations, pharmacokinetic data should be generated and used to develop pharmacokinetic models that make it possible to infer human internal doses (not just intake) from biomonitoring data and animal internal doses from administered doses.

In the case of PBDEs, integration of human data with animal data was challenging because intelligence and attention measures in humans do not have directly corresponding measurements in rodent models. Furthermore, the animal studies used different tests of learning and memory and, even when the same type of behavioral test was used, testing methods and data analysis often differed between studies. The committee found it helpful to focus its quantitative analysis on a specific measure of learning that was most consistently reported in the animal studies.

Pharmacokinetic and mechanistic data provided biological plausibility that the effects observed in animal studies may reflect similar hazards in humans. The committee found that mechanistic data were useful during the scoping and problem formulation phase of planning the systematic review to help determine what outcomes to focus on, as well as to determine how the animal and human evidence could be integrated.

The phthalate and the PBDE evaluations are both cases in which current toxicity-testing paradigms identify a hazard that is presumed to be of concern to humans but might not accurately predict exposures at which humans are affected. The development of pharmacokinetic data and models for extrapolation of data from animal studies or human biomonitoring data could facilitate the evaluation of an EAC's potential to cause health effects in humans at low doses.

1

Introduction

To safeguard public health, the US Environmental Protection Agency (EPA) must keep abreast of new scientific information and emerging technologies, so it can apply them to public-health protection and regulatory decision making. In the chemical assessment arena, the agency has dealt for decades with questions about what animal-testing data to use to predict human-health hazards, estimate dose-response relationships, protect susceptible subpopulations, and address uncertainties. As alternatives to traditional toxicity testing have emerged, the agency has faced additional decisions about incorporating nontraditional data into its chemical assessments and whether such tests can replace some traditional testing methods. Calls for more transparency in assessments that influence regulatory decisions have led to the agency's use of better-defined and -documented approaches for evaluating and integrating evidence. More recently, EPA has embraced recommendations that its Integrated Risk Information System (IRIS) program use approaches that are more systematic in their literature-based assessments. Some context about these challenges and how they led to the request for this report are briefly discussed below.

ADVANCES IN TOXICITY TESTING

Toxicity tests evaluate chemicals for their potential to cause cancer, birth defects, and other adverse health effects. Information from toxicity testing serves as an important basis for public-health protection and regulatory decisions concerning chemicals. Traditional test methods were developed incrementally over several decades and are conducted using laboratory animals. Researchers typically test chemicals at high doses in animals to induce toxicity and to identify a dose at which no adverse effect is observed. Uncertainty factors are then used to derive a human exposure level designed to be protective of human health. Assumptions are made that dose-response relationships from these studies can be extrapolated below the tested dose range and that the effects observed in animals are relevant to humans. Examples have shown, however, that these assumptions are not applicable to all chemicals (e.g., Maronpot et al. 2004; NRC 2004; Bracken et al. 2009), and debates often arise about how to account for differences in response between test animals and humans. Research directed at addressing those debates has focused on understanding the mechanisms of how chemicals cause adverse effects, with increasing emphasis being placed on evidence of biological effects (versus overt toxicity) as a predictor of adverse effects.

In the early 2000s, EPA and other federal agencies began to alter their testing strategies to focus on developing in vitro test methods that would allow less expensive and more rapid toxicity screening of large numbers of chemicals. The National Research Council (NRC) report *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC 2007) was influential in stimulating large-scale initiatives to determine how in vitro testing methods could be used to predict human toxicity, most notably the Tox21 program. The Tox21 program is a collaborative effort among EPA's National Center for Computational Toxicology, the National Toxicology Program, the National Institutes of Health National Center for Advancing Translational Sciences, and the US Food and Drug Administration. The program's work focuses on understanding chemically induced biological effects and developing predictive models of human biological response. An important contribution to the Tox21 program is the chemical screening results from EPA's Toxicity Forecaster (ToxCast). The ToxCast program has used hundreds of high-throughput

screening methods and computational toxicology approaches to evaluate thousands of chemicals. High-throughput assays evaluate specific types of bioactivity at the molecular or cellular level. Information about a particular biochemical interaction might help predict modulation of specific biological pathways that can lead to an adverse outcome. Thus, the number of biological targets for ascertaining the toxicity of a chemical has grown substantially.

EPA has used computational methods to estimate the human daily oral doses that would produce steady-state in vivo blood concentrations equivalent to chemical concentrations used in ToxCast assays. The estimates were compared with data on human oral exposures to identify chemicals for which the values overlap, and in vivo effects associated with this subset of chemicals (evaluated from the standard toxicological studies required for product registration) were compared with assay end point hits (Rotroff et al. 2010; Wetmore et al. 2012; Judson et al. 2014). One goal of these evaluations was to predict reproductive, developmental, and cancer end points (Judson et al. 2014).

ENDOCRINE ACTIVE CHEMICALS

Endocrine active chemicals (EACs), also referred to as endocrine disruptors, have the ability to modulate hormone function by mimicking, blocking, or otherwise altering activities of endogenous hormones. Because small alterations in hormone concentrations or activities, particularly during sensitive life stages, could have lasting and significant effects, environmental exposures to these types of chemicals are of particular concern. EPA was directed by Congress in the 1996 Food Quality Protection Act and the Safe Drinking Water Act amendments to begin screening pesticides for their potential to produce effects similar to those produced by estrogen in humans. The agency also was given authority to screen other chemicals and to evaluate other endocrine effects. In response to the congressional mandate, EPA created the Endocrine Disruptor Screening and Testing Committee, which developed a two-tiered screening approach. The first tier is used to screen chemicals for their potential to interact with the estrogen, androgen, or thyroid hormone systems, and EPA has worked for many years on developing and validating a battery of assays for these purposes. Because of developments associated with the Tox21 program, an initiative called "EDSP21" was begun in 2012 to validate and incorporate computational or in silico models and in vitro high-throughput screening methods into the program. The second tier of testing involves identifying adverse endocrine-related effects and quantitatively characterizing the dose-response relationship. Despite advancements and expansion of some testing protocols to include endocrine sensitive end points, concerns continue to be raised that traditional toxicity-testing practices might not include evaluation of end points relevant to endocrine disruptors (Vandenberg et al. 2012; EFSA 2013; Gore et al. 2015).

DEFINING LOW DOSE

Early in its deliberations, the committee considered several definitions of low dose that were developed by other National Academies committees, EPA, and the National Toxicology Program (NTP). For example, EPA (2013) adopted NTP's definition of a low-dose effect as "a biological change occurring in the range of typical human exposures or at doses lower than those typically used in standard testing protocols" (NTP 2001). An independent group that evaluated the scientific evidence on low-dose effects and nonmonotonic dose-response (NMDR) relationships for endocrine-disrupting chemicals in mammalian species adopted similar language (Melnick et al. 2002).[1]

For the purposes of this report, the committee looked for common ground to define low dose. Some definitions of low dose include language similar to that found in the EPA definition that obliquely defines a point of reference: "doses lower than those typically used in standard testing protocols." Those phrases help define an upper exposure range because the highest dose used in many guideline-driven animal tox-

[1]Low-dose effects were defined as "biologic changes that occur in the range of human exposures or at doses lower than those typically used in the standard testing paradigm of the US EPA for evaluating reproductive and developmental toxicity" (Melnick et al. 2002, p. 427).

icity studies is often determined by the maximum tolerated dose (MTD).[2] The lowest dose used in standard guideline-driven toxicity studies is frequently a log order lower than the MTD. The vast majority of relevant human exposures will fall below the levels that are used in regulatory toxicity tests. However, some animal studies use dose ranges that deviate from standard testing protocols, thereby blurring the distinction between doses used in animal studies and human exposures. Therefore, the committee considered the need for additional context when defining a high-dose point of reference based on animal toxicity studies.

The committee also considered elements found in other definitions. For example, some definitions of low dose, including that developed by EPA, contain the clause "a biological change occurring in the range of typical human exposures" without explicitly defining "typical" human exposures. Another issue that was debated at length related to the variety of terms used by the scientific community to describe dose (e.g., applied and internal dose) and exposure (e.g., exposure is an external measure, whereas dose is an internal measure) and how the terms might best be incorporated into the definition of low dose. The NRC report *Exposure Science in the 21st Century: A Vision and a Strategy* (NRC 2012) also commented on the inconsistent use of the terms *dose* and *exposure* and advocated primarily using the term exposure because it is more broadly applicable. Consistent with that recommendation, the committee used exposure rather than dose in its definition of low dose.

The committee also debated whether occupational exposures might represent a "typical" human exposure because they might be orders of magnitude higher than exposures that occur in the general population. Despite the differences inherent between environmental and occupational exposure (Semple 2005), the definition of low dose developed by the committee does not exclude occupational epidemiologic studies from the evidence stream for evaluating low-dose effects. That decision was deemed important because occupational studies could provide critical evidence regarding whether a chemical has endocrine-active effects. Additionally, occupational exposures are not necessarily higher than nonoccupational exposures for certain types of compounds and uses. Therefore, the committee's definition recognizes that a range of exposures from environmental to occupational might be relevant to the investigation of low-dose endocrine effects. However, the committee recognizes that, in trying to define low-dose endocrine effects, information on effects at environmental exposures might take precedence over or have higher priority than information on effects at higher exposures that might occur in occupational settings.

Ultimately, the committee defined low dose as *external or internal exposure that falls within the range estimated to occur in humans. Human exposure estimates may be based on environmental or biomarker measurements and/or computational models; ideally, the estimates account for toxicokinetic processes. If no human exposure estimates are available, low dose is defined on a case-by-case basis relative to an explicitly specified exposure in a particular context, such as "below the U.S. EPA Reference Dose (RfD) for chemical X" or "below the point of departure (NOAEL or BMDL) derived from the extended one-generation study for chemical X."* This definition acknowledges that a single definition cannot be used in all contexts, so no generalizations can be made.

The committee recognizes that some readers might not be comfortable with its definition because it lacks a bright line by which low dose can be defined. Defining such a bright line is not strictly based on science but rather encompasses policy decisions that were beyond the purview of the committee. Similarly, although the idea of a threshold or the shape of the dose-response relationship is important to the historical etiology of the concept of low dose (Chateauraynaud et al. 2014), the committee concluded that it was not necessary to address these ideas explicitly in its definition.

LOW-DOSE EFFECTS

Traditional animal testing was designed to identify effect levels by using a dosing regimen that includes a dose that elicits an overt response and at least two lower doses, one of which should be a dose

[2]An MTD is broadly defined as the highest dose of a chemical that can be administered to an animal without causing excessive toxicity.

where no effect is observed. In EPA chemical assessments, such studies are often used to identify a no-observed-adverse-effect level (NOAEL) or a lowest-observed-adverse-effect level (LOAEL), which is then used as a starting point for quantifying human reference levels. Concerns have been raised that low-dose effects of EACs are not being identified by traditional toxicity tests because the dosing regimen does not include doses within the range of human exposure or because some end points are not evaluated (e.g., Vandenberg et al. 2012; EFSA 2013; Gore et al. 2015). In some cases, scientists disagree about whether certain effects are adverse. Disagreements include the degree to which a biological change from normal is adverse; whether some effects are adaptive and have little effect on an organism; defining the degree to which functional capacity is so impaired as to be considered adverse; and uncertainties about the reliability and sensitivity of precursor effects to predict adverse outcomes.

NONMONOTONIC DOSE-RESPONSE CURVES

A conventional assumption of toxicology is that the dose-response relationship between a chemical and an adverse health effect will have a monotonic shape: that is, the slope of the curve does not change sign. NMDR curves change sign at one or more inflection points. In the case of chemicals that exhibit a U-shaped curve, effects are more prominent at low and high doses than they are at intermediate doses. Evidence for such curves could be missed if the doses tested are not low enough or not enough doses are tested to reveal such a shape. This situation is also true for identifying inverted U-shaped curves and more complex dose-response curves.

NMDR curves have been reported to be of particular concern for EACs (Vandenberg et al. 2012; Lagarde et al. 2015). Although NMDR relationships are often discussed in conjunction with low-dose effects, they are separate issues. In 2013, EPA issued a draft report, *State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures*, to evaluate the evidence on NMDR curves and to make judgments about EPA's toxicity-testing practices and the implications for its risk-assessment procedures. The draft concluded that exposure to EACs can result in NMDR curves for specific end points and that such curves were found more often in in vitro studies, at high doses, and for exposures of short duration. It also asserted that there was insufficient evidence that NMDR curves for adverse effects occur below traditional thresholds derived from toxicity testing. An NRC committee reviewed that report and found deficiencies in how the agency conducted its evaluation (NRC 2014a). Specifically, the committee noted that different literature-evaluation strategies were used to assess chemicals that affect the estrogen, androgen, and thyroid systems, and these differences weakened the agency's ability to draw firm conclusions. The committee recommended that a more structured and formal process be used to evaluate different evidence streams, in keeping with recommendations of other NRC reports (NRC 2011, 2014b). More recent reviews of the evidence on NMDRs have been conducted by ANSES (the French Agency for Food, Environmental and Occupational Health and Safety) and the European Food Safety Authority (EFSA) (Lagarde et al. 2015; Beausoleil et al. 2016).

Although the committee was not charged with specifically addressing NMDR curves, it notes that important considerations in evaluating such relationships is whether studies have sufficient statistical power or a broad enough dose range to identify whether an NMDR exists. Human epidemiologic data in particular might lack the power to identify an NMDR relationship. That opinion is supported by a recent EFSA report that reviewed the evidence of NMDRs for substances relevant to food safety. Beausoleil et al. (2016) developed screening criteria—an indication of an NMDR was reported in the study, at least three dose groups were used, and the relevant test substance was not tested in the form of a mixture—to evaluate the relevance and reliability of the evidence from in vivo, in vitro, and human epidemiologic studies. Data sets from studies deemed relevant and reliable were later evaluated using statistical software and then ranked on the basis of the strength of the evidence for an NMDR curve. None of the human data sets were considered suitable for NMDR analysis. Challenges with using human data were found in another study that developed qualitative methods to assess published in vivo, in vitro, and human epidemiologic studies for the presence of NMDRs (Lagarde et al. 2015).

SYSTEMATIC REVIEW

Two NRC reports (2011, 2014b) have made recommendations that EPA's IRIS program use more transparent and consistent methods for conducting its toxicological evaluations. Systematic review methods were identified as having the necessary elements to support those types of assessments. A systematic review uses explicit, prespecified methods to identify, select, assess, and summarize findings of similar but separate studies to answer a focused research question. The systematic review process is undertaken to identify all relevant studies on the agent of interest, to evaluate the studies identified, and to provide a qualitative and, where possible, a quantitative synthesis of the identified studies. Methods for conducting systematic reviews of the comparative effectiveness of clinical interventions are well established (e.g., Cochrane Collaboration [Higgins and Green 2011] and IOM [2011]), and the methods have been adapted (e.g., Woodruff and Sutton 2014; NTP 2015) and used to answer environmental health questions. EPA has also begun incorporating systematic review approaches into its chemical evaluations.

STATEMENT OF TASK

Faced with the challenges described above, EPA requested that the National Academies convene an ad hoc committee to develop a strategy for evaluating evidence of low-dose adverse human effects that act through an endocrine-mediated pathway. The verbatim statement of task is provided in Box 1-1. In response to this request, the National Academies convened the Committee on Endocrine-Related Low-Dose Toxicity, which prepared this report. Biographical information on the committee members is presented in Appendix A.

COMMITTEE'S APPROACH

The statement of task had several components that could be interpreted in a variety of ways. The following is the committee's interpretation of the task and how it approached and addressed each of the components. One objective of the task was to develop a strategy for EPA to evaluate evidence of low-dose adverse human health effects, with the aim of using the strategy to evaluate whether its toxicity-testing practices are adequate for identifying low-dose effects. The committee's strategy included the use of the systematic review method as an investigation and data analysis tool. This helped tie the strategy with other major elements of the task, namely the performance of systematic reviews for animal and human evidence streams for two chemicals. The other components of the task were performed in support of the strategy or as part of the systematic review process. Key definitions for terms used throughout the report are provided in Box 1-2.

The requirement that the committee perform systematic reviews of at least two chemicals appeared to be a request for a demonstration of how to perform such reviews and for consideration of how systematic review methods would fit in an overall strategy for evaluating low-dose effects. From a methodologic standpoint, two sets of guidelines for performing systematic reviews for environmental health assessments were considered by the committee: the Navigation Guide (Woodruff and Sutton 2014) and the National Toxicology Program's Office of Health Assessment and Translation (OHAT) method (Rooney et al. 2014; NTP 2015). The guidelines have similar methods, and the committee used both approaches to conduct its systematic reviews.

It was necessary to focus on EACs that have a robust database of human and animal studies because the task specified that the committee demonstrate the integration of animal and human evidence. The committee judged that the strongest evidence would likely be from developmental exposure, so preference was given to developmental outcomes. A number of potential chemicals were considered, including

bisphenol A, DDT, genistein, methoxychlor, parabens, perfluorooctanoic acid, phthalates, polybromindate diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs), 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), and triclosan.

A public workshop was held on February 3, 2016, to assist the committee with selecting the topics for its systematic reviews. The committee decided that three case examples would be the focus of the workshop, with the expectation that the discussions would be relevant to other EACs. The three chemicals (and end points) explored at the workshop were phthalates (male reproductive malformations), TCDD

BOX 1-1 Verbatim Statement of Task

An ad hoc committee under the auspices of the National Research Council (NRC) will develop a strategy for evaluating evidence of low-dose adverse human effects that act through an endocrine-mediated pathway. The study will include a scientific workshop to support the conduct of systematic reviews of human and animal toxicology data for two or more chemicals that affect the estrogen, androgen, or perhaps other endocrine systems. The workshop will seek to identify examples of relevant chemicals, populations/model systems, and end points of interest for further study using systematic review methods. Systematic reviews for these chemicals/populations/end points for human and animal data streams will be performed under the direction of the committee. The committee will evaluate the results of the systematic reviews, demonstrate how human and animal data streams can be integrated, determine whether the evidence supports a likely causal association, and evaluate the nature and relevance of the dose-response relationship(s). The committee will consider how to use adverse outcome pathway (AOP) or other mechanistic data, including high-throughput data and pharmacokinetic information, to elucidate under what circumstances human and animal data may be concordant or discordant. The results of the committee's evaluation of low-dose toxicity can be used to inform EPA on the adequacy of its current regulatory toxicity-testing practices.

BOX 1-2 Definitions of Terms Used in the Consensus Study Report

- **Adverse effect:** A biological change in an organism that results in an impairment of functional capacity, a decrease in the capacity to compensate for stress, or an increase in susceptibility to other influences (adapted from IPCS 2004).
- **Adverse outcome pathway:** A conceptual description of the sequence of causally linked events at various levels of biological organization.
- **Biomonitoring:** A method for assessing human exposure to chemicals by measuring the chemicals or their metabolites in human specimens, such as blood or urine (CDC 2009).
- **Low dose:** An external or internal exposure that falls within the range estimated to occur in humans. Human exposure estimates may be based on environmental or biomarker measurements (e.g., air monitoring, National Health and Nutrition Examination Survey) and/or computational models (e.g., physiologically based pharmacokinetic models, ExpoCast); ideally, the estimates account for toxicokinetic processes. If no human exposure estimates are available, low dose is defined on a case-by-case basis relative to an explicitly specified exposure in a particular context, such as "below the US EPA reference dose (RfD) for chemical X" or "below the point of departure (NOAEL or benchmark dose) derived from the extended one-generation study for chemical X."
- **Systematic review:** A "scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

(male reproductive effects), and bisphenol A (female reproductive effects). Experts on these chemicals and other EACs were invited to serve as panelists at the workshop to discuss the available data and issues that are relevant to performing a systematic review. Draft PECO statements[3] and a series of questions to consider provided the platform for the workshop discussions. The questions posed to the panelists were aimed at facilitating discussion about what end points had enough data on which to base a systematic review that would ultimately help answer questions about toxicity-testing practices. The PECO statements and the questions posed during the workshop are presented in Appendix B.

Informed by the deliberations of the February 2016 workshop, the review articles on EACs, and the expertise of the committee members, the committee decided to focus on two chemical classes and specific developmental outcomes: phthalate effects on male reproductive-tract development and PBDE effects on neurodevelopment. During the course of exploring the database on PBDEs, the committee discovered that a systematic review of the association between developmental exposures to PBDEs and human neurodevelopment was under way and soon to be completed (Lam et al. 2015). Because of this, PBDEs were removed from consideration as a case example at the workshop. However, after further consideration, the committee decided to conduct a case study of how an existing systematic review could be used by EPA. In this case example, the existing systematic review on PBDEs and human neurodevelopment was evaluated and updated by the committee and the results integrated with an independent systematic review of the animal evidence on PBDEs and neurodevelopment. Thus, the committee performed three independent systematic reviews and updated one existing systematic review. The approach used by the committee for its independent reviews was the OHAT method, whereas the authors of the existing review followed the Navigation Guide method. Table 1-1 summarizes the approaches used in the different reviews and illustrates that the methods are compatible. It also presents the approaches used to integrate the evidence.

After the systematic reviews were performed, the committee integrated the human and animal evidence using an approach developed by OHAT (NTP 2015). Mechanistic information was considered in this step to make determinations about the biological plausibility of the observed effects, to consider concordance or discordance in the evidence, and to draw conclusions about the potential hazard to humans. The nature and relevance of dose-response relationships in animals and humans were examined for the purposes of making a determination about whether observed effects are associated with low doses. Where possible, the implications of the results for the purpose of evaluating the adequacy of EPA's toxicity-testing practices were examined.

ORGANIZATION OF THE CONSENSUS STUDY REPORT

The committee's report is organized into a summary, five chapters, and six appendixes. Chapter 2 presents the committee's proposed strategy for evaluating evidence of low-dose adverse human effects that act through an endocrine-mediated pathway. Chapters 3 and 4 focus on the systematic reviews performed by the committee. Chapter 3 presents the evaluation of the animal and human evidence on phthalates and male reproductive-tract development and describes how the evidence was integrated to draw conclusions about associations and low-dose effects. Chapter 4 presents the results of the systematic reviews and assessments that were performed on PBDEs and developmental neurotoxicity. The lessons learned from performing the systematic reviews and further reflections on the statement of task are provided in Chapter 5.

Appendix A provides biosketches of the committee members. The other appendixes are provided as a PDF file available at https://www.nap.edu/catalog/24758.

[3]A PECO (Population, Exposure, Comparator, and Outcome) statement is a framework to clarify aspects of the review question. It guides the literature search, inclusion and exclusion criteria, the types of data to be considered, and the approach to synthesizing the evidence.

TABLE 1-1 Methods Used in the Systematic Reviews and Evidence Integration Presented in Chapters 3 and 4

Step	Phthalate (Animal)	Phthalate (Human)	PBDE (Animal)	PBDE (Human)
Systematic Reviews				
Approach	Independent review by the committee.	Independent review by the committee.	Independent review by the committee.	Independent review by Lam et al. (2015, 2016[a]); review and update of the review by the committee.
Peer Review and Registration of Protocol	National Academies peer-review process.	National Academies peer-review process.	National Academies peer-review process.	Original protocol registered in PROSPERO[b]; protocol to update the review underwent National Academies peer-review process.
Review of Existing Systematic Review	Not applicable	Not applicable	Not applicable	Preestablished criteria for a systematic review and ROBIS[c]
Literature Search	Planned and performed in consultation with a librarian.	Planned and performed in consultation with a librarian.	Planned and performed in consultation with a librarian.	Planned and performed in consultation with a librarian.
Literature Screening	DistillerSR[d]	DistillerSR	DistillerSR	DistillerSR
Data Extraction	HAWC[e]	HAWC	HAWC	DRAGON[f]
Risk-of-Bias Evaluation	OHAT[g]	OHAT	OHAT	Navigation Guide[h]
Data Analysis and Evidence Integration	OHAT	OHAT	OHAT	Navigation Guide
Confidence Rating and Level of Evidence Conclusions	OHAT (based on GRADE[i])	OHAT (based on GRADE)	OHAT (based on GRADE)	Navigation Guide (based on GRADE)
Integration of Evidence				
Hazard Identification Conclusions	OHAT	OHAT	OHAT	OHAT (Navigation Guide ratings were translated into equivalent OHAT ratings.)

[a]Lam et al. (2016).
[b]PROSPERO is an international prospective register of systematic reviews (https://www.crd.york.ac.uk/PROSPERO).
[c]ROBIS is a tool to assess risk of bias in systematic reviews (Whiting et al. 2016).
[d]Distiller SR is an online application designed specifically for the screening and data extraction phases of a systematic review (https://www.evidencepartners.com/).
[e]HAWC (Health Assessment Workspace Collaborative) is a Web-based interface application for warehousing data and creating visualizations (https://hawcproject.org).
[f] DRAGON is an online Access-based application designed for the data extraction phases of a systematic review (ICF International; https://www.icf.com/solutions-and-apps/dragon-online-tool-systematic-review).
[g]NTP (2015).
[h]Woodruff and Sutton (2014).
[i]GRADE (Grading of Recommendations, Assessment, Development and Evaluation) is a system for grading the quality of evidence in systematic reviews (Guyatt et al. 2011; Higgins and Green 2011).

REFERENCES

Beausoleil, C., A. Beronius, L. Bodin, B.G.H. Bokkers, P.E. Boon, M. Burger, Y. Cao, L. De Wit, A. Fischer, A. Hanberg, K. Leander, S. Litens-Karlsson, C. Rousselle, W. Slob, C. Varret, G. Wolterink, and J. Zilliacus. 2016. Review of non-monotonic dose-responses of substances for human risk assessment. EFSA Supporting Publication 13(5):1027E [online]. Available: https://www.efsa.europa.eu/en/supporting/pub/1027e [accessed May 3, 2016].

Bracken, M.B. 2009. Why animal studies are often poor predictors of human reaction to exposure. J. R. Soc. Med. 102(3):120-122.

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]. Available: http://www.cdc.gov/exposurereport/ [accessed June 15, 2016].

Chateauraynaud, F., J. Debaz, and M. Fintz. 2014. Chemical substances on the frontiers of health security: Metrological controversies over endocrine disruptors and low doses. Soc. Sci. Inform. 53(4):437-452.

EFSA (European Food Safety Authority) Scientific Committee. 2013. Scientific opinion on the hazard assessment of endocrine disruptors: Scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. EFSA J. 11(3):3132 [online]. Available: http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2013.3132/full [accessed June 15, 2016].

EPA (U.S. Environmental Protection Agency). 2013. State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures. U.S. Environmental Protection Agency. June 2013. Available: https://www.epa.gov/sites/production/files/2016-01/documents/nmdr.pdf [accessed May 15, 2016].

Gore, A.C., V.A. Chappell, S.E. Fenton, J.A. Flaws, A. Nadal, G.S. Prins, J. Toppari, and R.T. Zoeller. 2015. EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. Endocr. Rev. 36(6):E1-E150.

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

Higgins, J.P.T., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available: www.handbook.cochrane.org [accessed March 8, 2017].

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Review. Washington, DC: The National Academies Press.

IPCS (International Programme on Chemical Safety). 2004. IPCS Risk Assessment Terminology. Part 1: IPCS/OECD Key Generic Terms Used in Chemical Hazard/Risk Assessment. Part 2: IPCS Glossary of Key Exposure Assessment Terminology. Geneva: World Health Organization [online]. Available: http://www.inchem.org/documents/harmproj/harmproj/harmproj1.pdf [accessed November 11, 2015].

Judson, R., K. Houck, M. Martin, T. Knudsen, R.S. Thomas, N. Sipes, I. Shah, J. Wambaugh, and K. Crofton. 2014. In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast Programme. Basic Clin. Pharmacol. Toxicol. 115(1):69-76.

Lagarde, F., C. Beausoleil, S.M. Belcher, L.P. Belzunces, C. Emond, M. Guerbet, and C. Rousselle. 2015. Non-monotonic dose-response relationships and endocrine disruptors: A qualitative method of assessment. Environ. Health 14:13.

Lam, J., P. Sutton, J. McPartland, L.I. Davidson, N. Daniels, S. Sen, D. Axelrad, B. Lanphear, D. Bellinger, and T.J. Woodruff. 2015. Applying the Navigation Guide Systematic Review Methodology, Case Study #5. Association between Developmental Exposures to PBDEs and Human Neurodevelopment: A Systematic Review of the Evidence Protocol, April 2015 [online]. Available: http://www.crd.york.ac.uk/PROSPEROFILES/17890_PROTOCOL_20150322.pdf [accessed November 10, 2016].

Lam, J., B. Lanphear, D. Bellinger, D.A. Axelrad, J. McPartland, P. Sutton, L. Davidson, N. Daniels, S. Sen, and T. Woodruff. 2016. Systematic review and meta-analysis of association between PBDE exposure and IQ or ADHD. Draft report provided J. Lam (University of California, San Francisco) on July 1, 2016.

Maronpot, R.R., G. Flake, and J. Huff. 2004. Relevance of animal carcinogenesis findings to human cancer predictions and prevention. Toxicol. Pathol. 32(Suppl. 1):40-48.

Melnick, R., G. Lucier, M. Wolfe, R. Hall, G. Stancel, G. Prins, M. Gallo, M. Reuhl, S.M. Ho, T. Brown, J. Moore, J. Leakey, J. Haseman, and M. Kohn. 2002. Summary of the National Toxicology Program's report of the endocrine disruptors low-dose peer review. Environ. Health Perspect. 110(4):427-431.

NRC (National Research Council). 2004. Appendix A. Values and limitations of animal toxicity data. Pp. 159-167 in Intentional Human Dosing Studies for EPA Regulatory Purposes: Scientific and Technical Issues. Washington, DC: The National Academies Press.

NRC. 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: The National Academies Press.

NRC. 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2014a. Review of the Environmental Protection Agency's State-of-the Science Evaluation of Nonmontonic Dose-Response Relationships as They Apply to Endocrine Disruptors. Washington, DC: The National Academies Press.

NRC. 2014b. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2001. National Toxicology Program's Report of the Endocrine Disruptors Low Dose Peer Review. National Toxicology Progrm, U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC.

NTP. 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

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 science assessments. Environ. Health Perspect. 122(7):711-718.

Rotroff, D.M., B.A. Wetmore, D.J. Dix, S.S. Ferguson, H.J. Clewell, K.A. Houck, E.L. LeCluyse, M.E. Andersen, R.S. Judson, C.M. Smith, M.A. Sochaski, R.J. Kavlock, F. Boellmann, M.T. Martin, D.M. Reif, J.F. Wambaugh, and R.S. Thomas. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol. Sci. 117(2):348-358.

Semple, S. 2005. Assessing occupational and environmental exposure. Occup. Med. 55(6):419-424.

Vandenberg, L.N., T. Colborn, T.B. Hayes, J.J. Heindel, D.R. Jacobs, Jr., E.H. Lee, T. Shioda, A.M. Soto, F.S. vom Saal, W.V. Welshons, R.T. Zoeller, and J.P. Myers. 2012. Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. Endocr. Rev. 33(3):378-455.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell, III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

Woodruff, T.J., and P. Sutton. 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(10):1007-1014.

2

Strategy for Evaluating Low-Dose Effects

Approaches used by the US Environmental Protection Agency (EPA) for testing chemicals and for establishing health reference values have been modified over the past two decades in response to a growing awareness that some chemicals interact with endocrine hormone systems. Some changes made by EPA have been mandated legislatively. For example, as discussed in Chapter 1, the 1996 Food Quality Protection Act (FQPA) requires EPA to screen pesticides and other chemicals for their potential to produce estrogenic and other endocrine effects. Passage of the FQPA led to EPA's Endocrine Disruptor Screening Program and development of a series of in vitro and in vivo screening tests to identify chemicals that interact with the estrogen, androgen, or thyroid hormone systems (74 Fed. Reg. 54416 [2009]). Since the 1990s, EPA, the Organisation for Economic Cooperation and Development (OECD), and the National Toxicology Program (NTP) have modified toxicity-testing guidelines to improve their ability to detect effects that occur later in life after exposure to endocrine active chemicals (EACs) during sensitive windows of development. Toxicity-testing methods developed by OECD that can detect endocrine toxicity in mammals include the rodent two-generation reproduction study (TG 416), the extended one-generation reproductive toxicity study (TG 443), the rodent reproduction/developmental toxicity screening test (TG 421), the rodent chronic toxicity and oncogenicity studies (TG 451, TG 452, and TG 453), and the enhanced 28-day toxicity study (TG 407) (Bars et al. 2012). NTP has also updated its testing protocols: for example, by adding early life exposure to some cancer bioassays, adding endocrine outcomes, and extending follow-up when studying adverse effects on reproduction and development (Foster 2014).

Several aspects of EACs have prompted the need to assess the adequacy of traditional toxicity-testing strategies (see Box 2-1). Some EACs can mimic natural hormones, which affect the endocrine systems at low concentrations. Questions have been raised about how to use information about endocrine activity to understand potential health risks. For example, should a change in hormone concentrations be considered an adverse health effect? If such a change is not necessarily adverse, can it be used to predict an adverse outcome? If so, can one estimate the probability of an adverse outcome given an exposure?

BOX 2-1 Potential Limitations of Traditional Toxicity Testing for Evaluating Low-Dose Effects of Endocrine Active Chemicals

- Protocols might not include relevant windows of exposure or evaluation.
- Studies might not include relevant outcomes.
- Traditional animal models might be insensitive to certain effects seen in humans.
- Studies might not be conducted at environmentally relevant doses.
- Studies might lack the statistical power to detect effects at environmentally relevant doses.
- Dose selection might be inadequate to determine whether a nonmonotonic dose-response relationship exists.
- Single chemicals are typically evaluated.
- Genetically homogenous animal models might not reflect variability in human populations.

EACs also have the potential to cause long-lasting developmental effects. Organisms can be especially sensitive to EACs because hormones play a critical role during normal development. Dose and timing also can dramatically influence not only the magnitude of an effect but also the type of outcome observed or, in some cases, the direction of the effect (Ankley and Villeneuve 2015). For example, the synthetic estrogen diethylstilbestrol (DES) produces tumors in different tissues depending on what dose is administered and whether the exposure occurs prenatally or neonatally (Newbold and McLachlan 1982; Newbold 2004; Newbold et al. 2004). Thus, evaluation of EAC effects throughout an organism's lifespan is increasingly recognized as important, and some study designs that assess fertility, reproductive-tract malformations, or tumor incidence in animals have been modified. The modifications—such as dosing of pregnant animals throughout gestation, longer follow-up periods (including evaluation of several generations), assessment of multiple hormone-sensitive end points, and examination of multiple pups per litter—have improved sensitivity to detect endocrine effects (Blystone et al. 2010; Foster 2014). Awareness that EAC exposure during development can program tissues to respond differently to endogenous hormones or exogenous chemical challenges later in life (Newbold et al. 2004; Jenkins et al. 2007) or produce heritable modifications via epigenetic changes (Jefferson et al. 2013) is also growing.

Another concerning aspect of EACs is that effects from EAC exposure have been reported in humans and wildlife (Bernanke and Köhler 2009). Effects have been identified from a range of human exposures from environmental to deliberate administration of pharmaceutical agents. For example, DES administration to pregnant women provides a prime example of unexpected adverse effects, such as increased risk of breast cancer in mothers and various adverse outcomes in their offspring (Hoover et al. 2011; Reed and Fenton 2013).

One additional factor that has prompted increased scrutiny of EACs is the debate about whether thresholds exist for EAC effects. As reviewed by Hass et al. (2013), arguments in support of a threshold cite homeostatic mechanisms that are involved in endocrine regulation and the resiliency of higher order systems to adapt. Arguments against the presence of a threshold note that small fluctuations in endogenous hormones can affect regulation of a variety of biological processes (Hass et al. 2013). Thus, questions have been raised about whether the dose-selection practices used in traditional toxicity testing should be revised.

To protect public health and the environment, EPA and other agencies will need to work proactively to update testing methods as new science emerges. This chapter describes the committee's proposed strategy to assist EPA with the tasks of developing and revising testing practices in response to expanded knowledge about the potential for low-dose effects of EACs.

OVERVIEW OF THE COMMITTEE'S STRATEGY

The overall strategy envisioned by the committee for evaluating evidence of low-dose adverse human effects consists of three broad phases—surveillance, investigation and analysis, and actions (see Figure 2-1). The strategy recognizes that a toxicity testing program, no matter how sophisticated, cannot provide 100% assurance that all adverse effects will be identified and can be prevented. Even with pharmaceuticals, which are tested in human clinical trials, adverse effects are often not identified until after the drug has been marketed. Therefore, environmental chemicals, which are tested (if at all) in experimental systems only before use, require continued surveillance to protect public health given the expectation that false negatives will occasionally occur in testing.

Once a topic has been identified for additional investigation, the specific details of the investigation and analysis need to be planned so that they will support future agency actions. The results from the investigations and analyses are then used to select specific actions. In some cases, the only further action would be continued surveillance. In other cases where key uncertainties exist, further action could entail the generation of new data or models to address the uncertainties. If the results of an investigation suggest that adverse outcomes in humans are expected or might be occurring at low doses, the conclusions of previous toxicity testing or toxicity assessments for the chemicals that are under investigation might need to

be updated to reflect the new evidence. Additionally, such evidence might support updates to specific toxicity-testing or assessment practices to reduce the false-negative rate in the future.

The first two phases, *surveillance* and *investigation and analysis,* are described in more detail below. As noted in Figure 2-1, completion of each phase could involve one or more approaches. Although the descriptions of the components are presented sequentially in the report, there is no requirement that each approach be used or that a specific order be followed. The last phase, *actions,* involves policy decisions, which are outside of the committee's charge and therefore are not discussed in detail in this report.

SURVEILLANCE

In the context of this report, *surveillance* refers to the process for detecting signals (indications that an adverse outcome in a human population or animal model might be related to exposure to an EAC at low doses) by searching, retrieving, and evaluating *existing data*. Surveillance also refers to the process for monitoring the literature for methods that could be used for toxicity testing of EACs. Types of data that could be considered in an active surveillance program include human, animal, or mechanistic data.

Actively Monitor for New Data

A surveillance program for identifying low-dose effects should have a process for actively monitoring for new data to help ensure that effects will be identified and analyzed on a regular basis. Three broad categories—chemical-specific data, information that could lead to modifications of toxicity-testing methods and best practices for EACs, and information on endocrine-related effects in animals and humans—should be considered in the surveillance program. Relevant information to monitor might include scientific literature, various databases, nontraditional information sources, stakeholder input, and human exposure information. Although monitoring scientific literature and databases and human exposure information might provide the most valuable information, nontraditional information sources and stakeholder input have recently been highlighted as potential sources that could lead to valuable insights. Which sources are selected for surveillance will clearly depend on the problem under consideration and the resources available to the agency. Each source is discussed in more detail in the following sections.

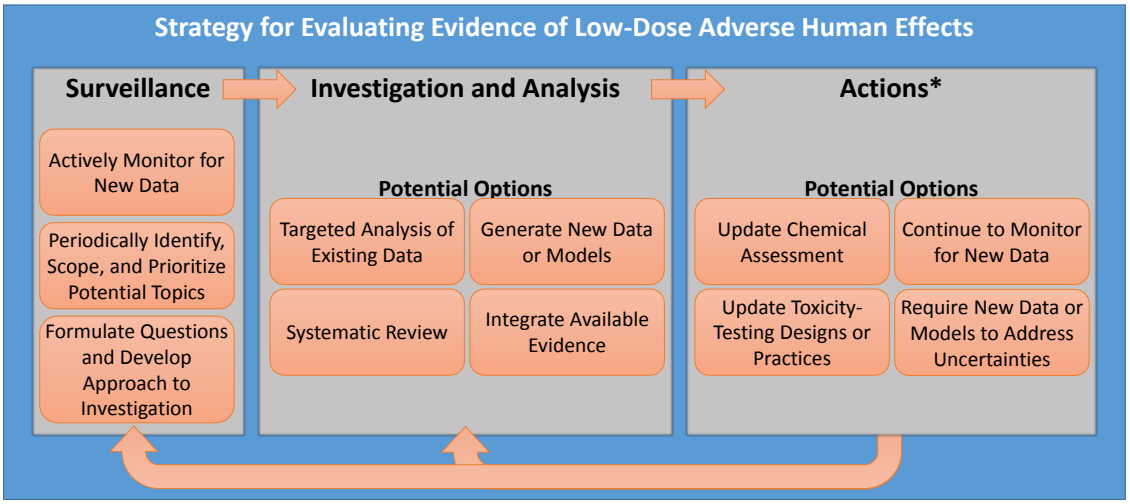


FIGURE 2-1 Strategy for evaluating evidence of adverse human effects from low-dose exposure to chemicals. The strategy includes three broad phases: surveillance, investigation and analysis, and actions. Each phase includes multiple options that may be employed alone or in combination. The order in which the options appear does not indicate a hierarchy or a sequence that should be followed. *Recommendations for this phase of the strategy were outside of the committee's charge.

Monitoring the Scientific Literature

EPA has many ongoing literature-review activities for evaluating chemical hazards, such as the chemical-specific evaluations of the Office of Research and Development, the Office of Pesticide Programs, the Office of Water, and the Office of Air and Radiation. Chemical-induced effects on endocrine function might be considered in those assessments. Although the assessments typically rely heavily on animal data, human data also contribute to these assessments. Whether human studies are concordant or discordant with the animal data is an important consideration.

In addition to chemical-specific data, the scientific literature should be monitored to identify relevant end points that might need to be included in toxicity testing or risk assessment or to identify other changes to toxicity-testing practices that might be needed to improve assessment of EAC effects. For example, several assessments conducted by EPA, NTP, and international institutions have investigated various methodological issues, including addition of mammary gland assessment in regulatory guideline protocols (Makris 2011); adequacy of rodent models for detecting hormonally related cancers (Thayer and Foster 2007); study design issues for developmental reproductive toxicity (Blystone et al. 2010); and the question of biological thresholds for EACs (Haas et al. 2013). Furthermore, literature reviews of diabetes, breast cancer, and other diseases have identified biological processes and specific chemicals that appear to be involved in their etiology (Rudel et al. 2007, 2011, 2014; Macon and Fenton 2013; Schwarzman et al. 2015; Auerbach et al. 2016; Smith et al. 2016; Bruner-Tran et al. 2017), and these reviews might reveal effects that need to be integrated into toxicity-testing methods.

The study of endocrine-related human diseases, such as prostate cancer and endometriosis, might also identify adverse effects that are not readily detected in rodent studies. For example, endometriosis is a human disease that is difficult to assess in animal models because most nonprimate mammals do not menstruate and consequently do not develop ectopic lesions, which are the pathological hallmarks of endometriosis (Bellofiore et al. 2017; Bruner-Tran et al. 2017). Concerns about the possible role of chemical exposure in the development of endometriosis might therefore be difficult to assess in an animal model (see Box 2-2). Agencies that are involved in health surveillance and that have health registries (e.g., Centers for Disease Control and Prevention, National Center for Health Statistics) could also be a source of information about relevant human diseases. Thus, by reviewing endocrine-related human diseases, one could ensure that signals are not being missed because they are not recapitulated in animal models. Given that there will always be species differences in response to chemical exposure, a surveillance system should always include monitoring epidemiologic literature or research.

The committee notes that automated methods are being developed for monitoring the literature. For example, methods or tools have been developed to extract drug-safety information from the published literature and electronic medical records (Shetty and Dalal et al. 2011; Wang et al. 2011; Duke et al. 2012; Gurulingappa et al. 2012; Avillach et al. 2013; Pontes et al. 2014; Winnenburg et al. 2015), and these approaches might be relevant for monitoring literature on EACs. Other publicly available tools have been developed for supporting systematic reviews (ICASR 2015), and these might assist with searching and retrieving data related to new toxicity-testing methods, outcomes, exposure assessment, and biomonitoring.

Assessing Scientific Data

In addition to the published peer-reviewed scientific literature, data are being generated and made available to the scientific community through various other venues. For example, as described in Chapter 1, EPA's ToxCast program has generated substantial data that can potentially provide toxicity and mechanistic information on a variety of chemicals. The testing is typically conducted in in vitro assays and uses a broad range of testing concentrations, including ones found in human biological samples; thus the testing could have relevance to environmental or low-dose exposure. The data are made available through

EPA's website.[1] Other information could come from databases developed to track mechanistic pathways for adverse outcomes.[2] These types of data could potentially help identify EACs, and their review might be considered important for a surveillance program.

Tracking Nontraditional Sources of Information

Some investigators have explored whether information extracted from such informal media sources as blogs and other various forms of social media could provide useful drug-safety surveillance data (Harpaz et al. 2012; Lardon et al. 2015; Nikfarjam et al. 2015; Sarker and Gonzalez 2015). Those efforts have proven challenging for drug surveillance programs because the language used to describe medical information in social media is often informal or descriptive (Nikfarjam et al. 2015). Applying those methods to surveillance of environmental chemical exposures could prove even more challenging because, unlike with drug surveillance, people are generally unaware of their environmental exposures. Despite the limitations, approaches for evaluating social media could be explored as a means to provide additional surveillance of potential health effects associated with chemical exposure.

Obtaining Stakeholder Input

A surveillance program could also include components from stakeholder input. EPA could evaluate recommendations or policy statements made by scientific societies or organizations or other academic groups. For example, several position papers have been published on the effect of EACs on human health (Diamanti-Kandarakis et al. 2009; Skakkebaek et al. 2011; Gore et al. 2015; Bennett et al. 2016). Another means of identifying a potential public-health problem is to consider input from advocacy groups,

BOX 2-2 Endometriosis: An Example in Which Traditional Toxicity Testing Might Be Inadequate to Evaluate Endocrine Active Chemicals

Human disease

- Endometriosis is characterized by the presence of an ectopic endometrium.
- This disease affects about 5-10% of women of reproductive age (Vercellini et al. 2014).
- The etiology is poorly understood and might involve developmental programming, immune modulation, and other mechanisms (Burney and Giudice 2012; Taylor et al. 2015).

Animal models

- Primates are the only models that menstruate and develop spontaneous endometriosis (Slayden 2016; Bruner-Tran et al. 2017).
- Although other animal models do not naturally develop endometriosis, rabbit and rodent models of endometriosis have been established by transplanting endometrium or uterine fragments from the same species (homologous models) or from humans (heterologous models) to ectopic sites (King et al. 2016). These models have not been used widely in toxicology.

Role of chemicals in endometriosis

- A nonstatistically significant doubling of risk for endometriosis was reported in women exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) following a factory explosion in Seveso, Italy, in 1976 (Eskenazi et al. 2002).
- TCDD induces endometriosis in primates (Rier et al. 1993) and a uterine phenotype in mice that mimics the reduced uterine progesterone responsiveness observed in women with endometriosis (Nayyar et al. 2007; Bruner-Tran et al. 2017).

[1]See https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data.
[2]See, for example, https:/aopwiki.org and https://www.effectopedia.org.

individual scientists, political leaders, chemical manufacturers, and other stakeholders. For example, NTP has a long-standing process that encourages nominations of chemicals of concern regarding human health (Heindel 1988). Similarly, EPA's Integrated Risk Information System (IRIS) program provides opportunities for EPA program and regional offices and other stakeholders to nominate chemicals for consideration. Nomination processes provide opportunities for stakeholder input and can help focus attention on societal concerns. Nomination processes are not without some limitations, however, including the possibility of reporting bias by the media, community action groups, or industry. Those groups might also be influenced by social activism or corporate product defense and might not be focused on longer-term public-health issues (Mihaylov and Perkins 2015; Zoller 2017). Guidance for engaging stakeholders has been provided in other National Academies reports (e.g., NRC 2009, 2014).

Monitoring Human Exposure Information

Biomonitoring data are an important information source that can help identify whether human exposure to EACs has occurred or changed over time (NRC 2006a, 2007, 2009, 2012) and are useful in defining low-dose exposures. They can also identify potential exposure sources and demographics of highly exposed groups. A previous NRC committee provided the following recommendations regarding the need to use biomonitoring data in surveillance programs: "Develop biomonitoring-based epidemiologic, toxicologic, and exposure-assessment investigations and public-health surveillance to interpret the risks posed by low-level exposure to environmental chemicals. Where possible, enhance existing exposure assessment, epidemiologic, and toxicologic studies with biomonitoring to improve the interpretation of results of such studies" (NRC 2006a, p. 9).

Box 2-3 illustrates how biomonitoring data from the US Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (NHANES) can help inform evaluations of EACs. Occupational exposure data might also provide information that could help define the range of human exposures to some chemicals.

In the absence of biomonitoring data, external exposure data can be used to estimate the range of potential human exposure to chemicals. Active air sampling devices have been used for several years to assess personal exposure to air pollutants, such as particulates, ozone, and polycyclic aromatic hydrocarbons (Geyh et al. 1999; Perera et al. 2003; Tsai et al. 2012; Oliveira et al. 2016). More recently, passive sampling devices have been developed to eliminate the need for cumbersome equipment that decreases compliance in human studies. Passive samplers are favored for personal monitoring because they are lighter and smaller and less likely to interfere with daily activities (NRC 1991). For example, silicone wristbands have been developed as personal passive samplers capable of sequestering polycyclic aromatic hydrocarbons, flame retardants, and pesticides as measures of an individual's external exposure (O'Connell et al. 2014; Donald et al. 2016; Hammel et al. 2016). Several funding agencies are supporting the development of sensor technologies for the 21st century, which includes wearable monitors that can be used in population studies to measure personal exposure in real time with high sensitivity and specificity and low cost.[3]

Another type of exposure data comes from efforts to measure chemicals in air, water, soil, and other environmental media (such as house dust), food, and consumer products, and extrapolate those measurements to human exposure. For example, EPA's Particulate Matter (PM) Supersites Program was established to obtain atmospheric measurements to address the research questions and scientific uncertainties about PM relationships between sources, receptors, exposures, and effects (Solomon and Sioutas 2008). A similar EPA program—the Clean Air Status and Trends Network (CASTNET)—is a national monitoring network established to assess trends in pollutant concentrations, atmospheric deposition, and ecological effects due to changes in air-pollutant emissions (Puchalski et al. 2015). Other air-monitoring programs have evaluated changes in urban ozone concentrations that have occurred over several decades (Sather and Cavender 2016).

[3]See https://www.sbir.gov/Sensor-technology-for-the-21st-century.

BOX 2-3 NHANES: A Platform for Surveillance

Since the National Health and Nutrition Examination Survey (NHANES) III (1988-1994) was conducted, biomonitoring has been expanded to include biomarkers of selected pesticides, phthalates, and volatile organic compounds (CDC 2009; Sobus et al. 2015). As of 2015, 265 chemical biomarkers—including ones for some brominated flame retardants, dioxins and furans, pesticides, metals, perfluorinated compounds, phthalates, and polychlorinated biphenyls—are assessed (CDC 2015). The survey also collects data on health end points and demographics and is designed to be representative of the US population.

NHANES biomonitoring data have been used to identify temporal trends in chemical exposures and can help define low-dose ranges. For example, Hartle et al. (2016) used 24-h dietary recall data and urinary samples to assess the association between consumption of canned foods and beverages and biomarkers of exposure to bisphenol A in a subset of the NHANES population that was 6 years of age and older to understand human exposure. NHANES studies can also support hypothesis generation related to health outcomes. For example, some studies have evaluated associations between urinary or blood biomarkers and hormone function, such as research that investigated links between urinary organophosphate insecticide concentrations and serum testosterone and estradiol concentrations in adult men (Omoike et al. 2015). Other studies have investigated links between serum perfluoroalkyl concentrations and serum testosterone, thyroid stimulating hormone, free and total triiodothyronine, and thyroxine levels in 12- to 80-year-old males and females (Lewis et al. 2015). Data derived from NHANES studies have also supported the development of computational dosimetry models, such as reverse toxicokinetic models, that can link chemical biomarker measurements to exposure levels (Tan et al. 2012; Sobus et al. 2015). NHANES data are also used to compare biomarker measurements to model-predicted biomarker estimates.

Biomonitoring data are not without limitations, however. NHANES biomarker data are determined for blood or urine samples that have been collected from volunteers at a single point in time. Single time-point measurement of a chemical biomarker might not accurately predict average or peak exposures (Aylward et al. 2013, 2014; Bradman et al. 2013) and might miss exposures that occur during pregnancy, fetal development, or other life stages.

Surveillance of Endocrine-Mediated Drug Reactions Using Pharmacovigilance

The US Food and Drug Administration (FDA) has developed pharmacovigilance[4] programs to monitor for adverse drug reactions (ADRs). Such programs are needed because human clinical trials use relatively small sample sizes; have shorter durations of exposure than usually occur; often lack diversity among study participants; and do not include pediatric and other susceptible subpopulations (McMahon et al. 2015). Those design features limit the ability of clinical trials to detect rare (1:1,000 to 1:10,000) ADRs that occur at therapeutic drug doses (Schotland et al. 2016).

Pharmacovigilance programs often consist of premarketing surveillance that identifies ADRs during preclinical screening and clinical trials and postmarketing surveillance in which data are accumulated throughout a drug's market life (Ibrahim et al. 2016). Voluntary reporting systems have historically served as the primary data collection system for postmarketing pharmacovigilance. Such passive surveillance systems primarily rely on the collection of reports of suspected ADRs from health care professionals, consumers, and pharmaceutical companies. Today, FDA has established the Office of Surveillance and Epidemiology to help coordinate its efforts in postmarket drug safety surveillance (see Box 2-4).

[4]Pharmacovigilance is defined by the World Health Organization (WHO) as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem" (WHO 2002, p. 7). Pharmacovigilance includes postmarketing safety surveillance activities to detect events that were not seen in a clinical trial ("safety signal generation").

BOX 2-4 Overview of FDA's Efforts in Postmarket Drug Safety Surveillance

The FDA Adverse Event Reporting System (FAERS) and the FDA-CDC Vaccine Adverse Event Reporting System (VAERS) represent examples of government-led voluntary reporting systems. The computerized FAERS database receives more than 500,000 new safety reports per year (Ibrahim et al. 2016). FAERS data are used to support regulatory actions, such as drug label updates and market withdrawals (Hauben et al. 2007; Tatonetti et al. 2012; Ibrahim et al. 2016). The FAERS database can also be used to investigate emerging drug safety issues and to inform hypotheses about plausible causes of ADRs (Sarkar et al. 2011; Ibrahim et al. 2016).

There are several important limitations of the FAERS database, including under-reporting of ADRs that results in underestimates of the prevalence of drug-ADR associations and uncertainty that a given ADR is causally related to the drug exposure (Ibrahim et al. 2016). Such limitations have led to the increased use of electronic health records as an additional source of pharmacovigilance data (Trifirò et al. 2009; Li et al. 2014). FDA has recently completed its Mini-Sentinel pilot project that developed methods, tools, resources, policies, and procedures to facilitate the use of routinely collected electronic health care data to perform active surveillance of the safety of marketed medical products (Platt et al. 2012; Gagne et al. 2016).

Some drugs have been shown to demonstrate endocrine activity (see, for example, Friedman et al. 2009). In fact, the Institute of Medicine recommended that FDA "engage the pharmaceutical industry and scientific community in postmarketing studies or clinical trials for hormonally active prescription drugs for which the potential impact on breast cancer risk has not been well characterized" (IOM 2012, p. 21). Such an approach is also relevant for other endocrine-related effects. The committee concludes that methods that have been developed for pharmacovigilance programs might be adapted or help inform an EAC surveillance program. The committee notes, however, that EAC surveillance programs that are based solely on voluntary reporting might be limited because people often cannot self-report environmental exposures.

Periodically Identify, Scope, and Prioritize Topics

Active surveillance will likely result in the identification of information on chemical exposures, outcomes, and advances in toxicity testing that will need to be reviewed. Accordingly, the next step in the committee's strategy (see Figure 2-1) is periodic review of information and identification of topics that might need to be pursued further. That effort could involve a scoping exercise in which EPA would survey the literature and other information to determine the extent, range, and nature of information available on the topic, to identify data gaps, and to consider whether additional research might be needed. The scoping step would assist EPA in setting priorities for topics that deserve further study. Decisions to pursue a topic could be influenced by a number of factors, including the size of the population at risk, public-health significance, and available resources.

Formulate Questions to Address and Develop an Approach for the Investigation

Once a topic is selected for further analysis, the next step is to formulate the questions to address and develop an approach to the investigation, which will involve consideration of the scientific evidence, expert judgment, and relevant stakeholder perspectives. Questions and approaches are often targeted depending on the topic under investigation and potential actions that could be taken by the agency. For example, EPA might consider whether a new outcome measure should be included in regulatory toxicity tests. In that case, questions that might be posed include the following:

- Is the outcome an emerging concern?
- Are appropriate assays available or could they be developed?

- Is the outcome a more sensitive measure of adverse toxicity than currently used outcomes?
- Is the outcome of interest a reproducible effect after chemical exposure?
- Will including the new outcome in regulatory testing improve hazard identification—that is, improve sensitivity?

Box 2-5 provides a retrospective example of how the answers to these questions helped establish the measurement of anogenital distance as an outcome measure in regulatory toxicity tests.

Once the questions have been identified, the next step is to formulate the approach to the investigation and determine the types of data and analyses that are needed to answer the questions and provide the basis for agency actions. Figure 2-1 shows four types of investigation and analysis that could be considered: generation of new data or models to fill data gaps, targeted analysis of existing data, systematic review, and integration of the available evidence. The types of investigations or analyses might not be mutually exclusive because several related investigations or analyses might be needed to address the questions adequately. And, as noted above, they could be influenced by a number of factors, including the size of the population at potential risk, public-health significance, and available resources.

BOX 2-5 Anogenital Distance and Its Addition to Regulatory Toxicity Tests

This example illustrates the types of questions that might be asked to address whether an outcome measure—anogenital distance (AGD)—should be included in regulatory toxicity tests. This example does not include all questions or issues that might have been considered in this particular decision.

1. *Is the outcome an emerging concern?* AGD is sexually dimorphic in many mammals; males have longer AGD than females do. Reduced AGD is considered a sensitive indicator of reduced fetal androgen (Liu et al. 2014) during the male programming window. Studies that explore chemical effects on AGD in animals date back at least 50 years (Revesz et al. 1960). Swan et al. (2005) were among the first to examine whether exposure to an EAC could result in altered AGD in humans.
2. *Are appropriate assays available?* Measurement of AGD is relatively straightforward in animals and is defined as the distance from the genital tubercle to the anus. Similar methods for the measurement of AGD in humans have been developed (Salazar-Martinez et al. 2004; Sathyanarayana et al. 2015).
3. *Is the outcome a more sensitive measure of adverse toxicity than currently used outcomes?* AGD is a more sensitive measure for reduced (fetal) androgen signaling than traditional outcomes such as hypospadias and cryptorchidism (Saillenfait et al. 2008; Kim et al. 2010).
4. *Is the outcome of interest a reproducible effect after chemical exposure?* Like other anthropometric measurements, errors in the measurement of the AGD can occur. Accurate AGD measurements depend on the identification of distinct anatomical landmarks. Measurement of AGD by a single trained examiner is preferred because inter-rater variability is often larger than intra-rater variability. A retrospective analysis of 43 multigeneration studies (16 in Wistar rats and 27 in Sprague-Dawley rats) conducted according to the latest version of the test guidelines indicated that measurement of AGD had a coefficient of variance of 25-50% (Marty et al. 2009). In humans, inter-rater and intra-rater reliability and thus reproducibility of AGD measurements can be very good if appropriate methods are followed that include standardized training and monitoring of measurements (Sathyanarayana et al. 2015).
5. *Will including the new outcome in regulatory testing improve hazard identification?* Scientific consensus was reached that AGD should be added to testing methods. For example, EPA revised OPPTS 870.3800/OECD 416 (Reproduction and Fertility Effects Test) to include AGD measurement in F2 offspring if triggered by a change in sex ratio or age at puberty onset (Marty et al. 2009). The measurement of AGD between postnatal days (PND) 0 and PND 4 was also added as a required outcome measure in OECD 421 (Reproduction/Developmental Toxicity Screening Test), OECD 422 (Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test), and OECD 443 (Extended One-Generation Reproductive Toxicity Study) tests (Beekhuijzen et al. 2016).

INVESTIGATION AND ANALYSIS

Targeted Analysis of Existing Data

In some cases, a targeted analysis (or reanalysis) of existing data might support agency actions. A targeted analysis might be used when different test systems, methods, analytical approaches, and other experimental design features differ between studies and make the results difficult to compare. For example, outcome measures that are analyzed as continuous variables (such as changes in mean response) in some studies are not directly comparable to outcome measures that are analyzed as dichotomous variables (such as numbers of individuals beyond a specified cut point) in other studies. Another example is the summary of dose-response data in terms of pair-wise significance. In that case, the need for additional analyses could arise when there is a finding of a no-observed-adverse-effect level (NOAEL) in one study and a lowest-observed-adverse-effect level (LOAEL) in another study at the same dose. A targeted analysis of these seemingly discordant data might be able to strengthen the interpretation of the evidence. Specifically, an analysis of those data should consider whether they are statistically significantly different from each other. Additionally, trend analyses might also be useful as an alternative to pair-wise comparisons, because trend tests might have greater statistical power. Other contextual factors—including experimental design and conduct, mechanistic data, and prior evidence—should also be considered (Goodman 2016). Consideration of those factors will help reduce the tendency of some regulators to incorrectly perceive a NOAEL as a threshold (Scholze and Kortenkamp 2007) and reduce reliance on the use of a *p*-value (most often set at 0.05) as a bright line in evaluating whether an effect has occurred (Goodman 2016).

It is also common that studies from different investigators will not be performed at the same doses, and this makes it difficult to compare studies and evaluate consistency. To interpolate between dose groups, one can typically use parametric or nonparametric curve-fitting approaches. For epidemiological data, replacing pair-wise comparisons with regression analyses allows comparison of regression coefficients that can address this issue. For both toxicological and epidemiological data, the benchmark dose (BMD) approach can be well suited for comparing evidence of adverse effects among both toxicologic and epidemiologic studies. The BMD is the estimated dose associated with a specific level of response, called the *benchmark response* (BMR), along with its confidence interval. The use of a common BMR forces a transparent definition of the size of a biologically significant effect that is common among the studies being compared, and the resulting confidence intervals can be compared to evaluate study consistency or inconsistency. That approach has been extended to end points that are not strictly identical by using categorical regression, in which disparate end points are grouped into "bins" of severity categories (EPA 2000). The approach then estimates the BMD associated with a specific severity. Categorical regression, however, requires judgment to determine which end points and magnitudes of effect are to be grouped together at each level of severity.

There are cases in which data might be reanalyzed to better account for uncertainties. For example, in its report *Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment*, NRC (2006b) recommended conducting a quantitative uncertainty analysis as part of EPA's dose-response assessment. There have been numerous other examples in the literature in which data have been reanalyzed to address previously unaddressed uncertainties, particularly at low doses (see, for example, Subramaniam et al. 2007; Crump et al. 2016).

Systematic Review

Systematic reviews provide a method for evaluating evidence in a transparent and consistent manner that reduces bias. As described in several National Academies reports (NRC 2011, 2014), systematic review provides a rigorous approach to evaluating evidence that, although not removing the role of expert judgment, aims to make such judgments more transparent and less susceptible to biases. Guidelines for performing systematic reviews relevant to environmental or public-health assessments include the Navi-

gation Guide (Woodruff and Sutton 2014) and NTP's Office of Health Assessment and Translation (OHAT) approach (Rooney et al. 2014; NTP 2015).

A systematic review is guided by a study question that should be carefully crafted to address the problem under consideration. If the question is too broad in scope, the studies included in the analysis might be too heterogeneous for effectively integrating and drawing a conclusion. On the other hand, if it is too narrow in scope, the results are less generalizable and might not be relevant to the underlying public-health concern. The methods for identifying, screening, and analyzing the scientific literature are planned in advance to ensure that the evidence is selected and evaluated in an objective and consistent manner. The committee notes that the systematic review method alone does not lend itself to answering a question about whether a specific EAC has low-dose adverse effects. At a minimum, answering that question requires the completion of hazard identification and dose-response assessment for the effects of interest.

Generate New Data or Models

A focused research question could be addressed by generating new data or models. In some cases, that activity could involve conducting animal toxicity studies, pharmacokinetic studies, or epidemiologic investigations. At other times, new in vitro data could shed light on the mechanisms involved in an observed response. Other types of research questions might be answered through computational model development. For example, physiologically based pharmacokinetic (PBPK) or other quantitative dosimetry models for a chemical of interest could be developed to compare dose-response relationships in human studies with those in rodent studies. Quantitative dosimetry models could also help define the relationship between external dose and internal tissue concentrations and could be used to support cross-species and route-to-route extrapolations. Other types of dosimetry models might be needed to facilitate interpretation of in vitro data. For example, reverse-toxicokinetic modeling and in vitro–in vivo extrapolations are used to compare in vitro data with estimated or measured human exposure data (Wetmore et al. 2012; Yoon et al. 2012).

Questions about the coherence of findings in rodents and humans might be addressed through evaluation of mechanistic data. One way of improving our understanding of mechanistic data involves the development of conceptual models that facilitate the organization of information about biological interactions and toxicity mechanisms. A conceptual model can reflect the initial interactions of a chemical with the biological system and the resulting events that can lead to a specific adverse outcome (Ankley et al. 2010). Tools are being developed that can predict associations between key events and thus lead to the development of quantitative or computationally predicted adverse outcome pathways (qAOPs or cpAOPs) (Bell et al. 2016; Connolly et al. 2017). The committee notes that endocrine-related effects probably do not result solely from one isolated or linear pathway but involve multiple pathways or networks, and disease manifestation likely involves multiple components or stressors (NASEM 2017).

Determining the types of new data and models that are needed should be tailored to the research questions, and specific recommendations are beyond the scope of this report. Nevertheless, Box 2-6 provides an example of generating data to address uncertainty.

Integrate Available Evidence

This committee—as have other National Academies committees before it—emphasizes the need for evidence integration to be both transparent and standardized in its approach. Thus far, evidence integration has focused on the purpose of hazard identification: that is, determining whether a causal relationship exists. Causal frameworks, such as those developed by the International Agency for Research on Cancer, NTP, and EPA, can be adopted or adapted to provide transparency and consistency in conducting causal evaluations. See previous National Academies reports for additional guidance on such approaches (NRC 2014; NASEM 2015, 2017).

BOX 2-6 Bisphenol A: Generating Data to Address Uncertainty

This example illustrates how uncertainty or data gaps can be addressed by generating new data and models.

Data gap: Uncertainties in the human pharmacokinetics of bisphenol A (BPA) after oral exposure. Specifically, some biomonitoring studies have reported serum concentrations of unconjugated BPA in the 1-10 nM range, which is similar to the range where some in vitro and in vivo studies have reported significant biological effects (Vandenberg et al. 2013).

Approach: To address this uncertainty, the National Institutes of Health conducted a human pharmacokinetic study using a single oral administration of deuterated BPA, which can be distinguished from background, and more sensitive analytical methods (LOD <10 pM) (Thayer et al. 2015). The study concluded that unconjugated BPA comprised less than 1% of total BPA in the serum, with elimination largely complete 24 hours after oral administration. Using those data with a new pharmacokinetic model (Yang et al. 2015) suggested that peak serum concentrations in the general population are likely to be about 5-20 pM for daily dietary intakes of up to 0.5 μg/kg-day, a range that is consistent with estimates based on other methods (Teeguarden et al. 2013, 2015).

Impact: Such concentrations are well below those studied in most in vitro and in vivo experimental studies, so most studies reporting "low dose" effects of BPA do not directly inform whether BPA can cause effects at current human exposure levels.

The committee recognizes, however, that for addressing low-dose adverse effects of chemical exposure, the question will often be more explicitly quantitative: that is, it specifically concerns the nature of the dose-response relationship at low doses. Although the evidence of causality is still important, the problem is that the causal evidence might include studies that exclusively include high exposures, such as experimental doses near the maximum tolerated dose. High-dose data alone are usually not useful for making inferences about response to exposures at low doses because of uncertainties in the shape of dose-response curves below the range of observation.

Therefore, it might be necessary to integrate the subset of evidence that includes low-dose toxicity data separately. Although not very informative for causal inferences, environmental exposure data, such as biomonitoring data, might nevertheless be useful for defining what subset of the data can be considered as low dose (see discussion earlier in this chapter). Some additional considerations include the following:

- In vitro–in vivo extrapolation or reverse toxicokinetics can help to determine what in vitro mechanistic data could be considered low-dose data.
- Incorporation of modeled dose-response relationships, including BMD estimates.
- Addressing toxicokinetic and toxicodynamic differences between species and populations.
- Biological plausibility given mechanistic data.
- Co-exposures that might act on the same end point.

Data integration can also be used to consider questions that are not about specific EACs but are broader, such as whether a new end point or new exposure or assessment window is relevant to determining low-dose effects. As noted earlier, some end points have been added to regulatory testing protocols in response to growing evidence that they are indicators of toxicity, and the duration of some tests has been extended to capture effects that might occur later in life. Signals identified during the surveillance step that have those types of implications about toxicity testing could be evaluated by integrating the available evidence. One example of such a signal is the growing concern about evaluations of mammary gland toxicity. Makris (2011) evaluated how the effects of environmental chemicals on the mammary gland are

assessed in guideline studies of EPA, OECD, and NTP and made a number of recommendations for enhancing how the end point is assessed. She identified data gaps, issues, and challenges and noted that "to address these issues, a paradigm shift would be needed for the evaluation of [mammary gland] in guideline studies" (Makris 2011, p. 1050). Challenges identified with implementing such a shift were issues of species and strain sensitivity, the timing of exposure, and when the end point is evaluated. Evidence integration could help address such issues.

ACTIONS

The remaining step in the committee's strategy is to select the types of actions that are needed. As shown in Figure 2-1, several types of actions could result, including the need to update chemical assessments, to continue to monitor for new data, to require new data or models to reduce uncertainties, or to update toxicity-testing designs and practices. The type of actions that EPA takes could be influenced by a number of factors, including the size of the population at risk, the public-health significance of the investigation, and available resources. Specific recommendations on exactly what actions to take are beyond the scope of this report.

FINDINGS AND RECOMMENDATIONS

- To ensure adequate understanding of hazards and to inform regulatory decision making, EPA needs a general strategy for ongoing evaluation of evidence of low-dose effects from exposure to EACs. The committee proposes a strategy involving three phases: surveillance, investigation and analysis, and actions. EPA is already conducting many activities consistent with the proposed strategy, though not necessarily in the specific context of assessing low-dose exposure to EACs.

  Recommendation: EPA should develop an active surveillance program focused specifically on low-dose exposures to EACs. This program could include regularly monitoring published research and other information sources, gathering input from stakeholders, and collecting human exposure information. It might also involve data collection in collaboration with other agencies and outside parties. The surveillance program should periodically identify, scope, and prioritize potential areas of focus related to low-dose effects, such as particular chemicals and end points. Some approaches discussed in this chapter will require methods and tool development, such as automated methods for monitoring the literature.

  Recommendation: After a topic is selected for further evaluation, the agency should plan its investigation by identifying key questions to be addressed and determining the types of data and analyses needed to answer the questions and to support future agency actions.

- The four main approaches for investigation and analysis are targeted analysis of existing data, systematic review, generation of new data or models, and integration of evidence. The types of analyses used to investigate the questions are not mutually exclusive, and several approaches might be needed to address the questions adequately. Integration of evidence for low-dose adverse human effects of EACs involves consideration of both hazard identification and dose response.

  Recommendation: Environmental exposure data should be used, if available, to define what subset of the data should be considered as low dose.

- A robust strategy will provide the agency with a range of options to address questions of concern.

  Recommendation: The specific approaches and tools used to implement the strategy to address issues related to low-dose endocrine effects will need to be considered on a case-by-case basis and should be guided by the specific questions under study.

- The proposed strategy in this chapter will facilitate more regular consideration of the adequacy of toxicity testing. However, the agency will also be faced with questions about the amount of evidence needed to change traditional test methods, and these questions might be more appropriately addressed through policy decisions.

REFERENCES

Ankley, G.T., and D.L. Villeneuve. 2015. Temporal changes in biological responses and uncertainty in assessing risks of endocrine-disrupting chemicals: Insights from intensive time-course studies with fish. Toxicol. Sci. 144(2):259-275.

Ankley, G.T., R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrano, J.E. Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29(3):730-741.

Auerbach, S., D. Filer, D. Reif, V. Walker, A.C. Holloway, J. Schlezinger, S. Srinivasan, D. Svoboda, R. Judson, J.R. Bucher, and K.A. Thayer. 2016. Prioritizing environmental chemicals for obesity and diabetes outcomes research: A screening approach using ToxCast™ high-throughput data. Environ. Health Perspect. 124(8):1141-1154.

Avillach, P., J.C. Dufour, G. Diallo, F. Salvo, M. Joubert, F. Thiessard, F. Mougin, G. Trifirò, A. Fourrier-Réglat, A. Pariente, and M. Fieschi. 2013. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: A contribution from the EU-ADR project. J. Am. Med. Inform. Assoc. 20(3):446-452.

Aylward, L.L., R. Kirman, R. Schoeny, C.J. Portier, and S.M. Hays. 2013. Evaluation of biomonitoring data from the CDC National Exposure Report in a risk assessment context: Perspectives across chemicals. Environ. Health Perspect. 121(3):287-294.

Aylward, L.L., S.M. Hays, R. Smolders, H.M. Koch, J. Cocker, K. Jones, N. Warren, L. Levy, and R. Bevan. 2014. Sources of variability in biomarker concentrations. J. Toxicol. Environ. Health B Crit. Rev. 17(1):45-61.

Bars, R., I. Fegert, M. Gross, D. Lewis, L. Weltje, A. Weyers, J.R. Wheeler, and M. Galay-Burgos. 2012. Risk assessment of endocrine active chemicals: identifyingchemicals of regulatory concern. Regul. Toxicol. Pharmacol. 64(1):143-154.

Beekhuijzen, M., F. van Otterdijk, W. Wieland, M. van Tuyl, R.P. Rijcken, B. Peter, and H. Emmen. 2016. Update of OECD DART guidelines with endocrine disruptor relevant endpoints: Practical considerations. Reprod. Toxicol. 64:64-71.

Bell, S.M., M.M. Angrish, C.E. Wood, and S.W. Edwards. 2016. Integrating publicly available data to generate computationally predicted adverse outcome pathways for fatty liver. Toxicol. Sci. 150(2):510-520.

Bellofiore, N., S.J. Ellery, J. Mamrot, D.W. Walker, P. Temple-Smith, and H. Dickinson. 2017. First evidence of a menstruating rodent: The spiny mouse (*Acomys cahirinus*). Am. J. Obstet. Gynecol. 216(1):40e1-40e11.

Bennett, D., D.C. Bellinger, L.S. Birnbaum, A. Bradman, A. Chen, D.A. Cory-Slechta, S.M. Engel, M.D. Fallin, A. Halladay, R. Hauser, I. Hertz-Picciotto, C.F. Kwiatkowski, B.P. Lanphear, E. Marquez, M. Marty, J. McPartland, C.J. Newschaffer, D. Payne-Sturges, H.B. Patisaul, F.P. Perera, B. Ritz, J. Sass, S.L. Schantz, T.F. Webster, R.M. Whyatt, T.J. Woodruff, R.T. Zoeller, L. Anderko, C. Campbell, J.A. Conry, N. DeNicola, R.M. Gould, D. Hirtz, K. Huffling, P.J. Landrigan, A. Lavin, M. Miller, M.A. Mitchell, L. Rubin, T. Schettler, H.L. Tran, A. Acosta, C. Brody, E. Miller, P. Miller, M. Swanson, and N.O. Witherspoon. 2016. Project TENDR: Targeting environmental neuro-developmental risks the TENDR consensus statement. Environ. Health Perspect. 124(7):A118-A122.

Bernanke, J., and H.R. Köhler. 2009. The impact of environmental chemicals on wildlife vertebrates. Rev. Environ. Contam. Toxicol. 198:1-47.

Blystone, C.R., G.E. Kissling, J.B. Bishop, R.E. Chapin, G.W. Wolfe, and P.M. Foster. 2010. Determination of the di-(2-ethylhexyl) phthalate NOAEL for reproductive development in the rat: Importance of the retention of extra animals to adulthood. Toxicol. Sci. 116(2):640-646.

Bradman, A., K. Kogut, E.A. Eisen, N.P. Jewell, L. Quirós-Alcalá, R. Castorina, J. Chevrier, N.T. Holland, D.B. Barr, G. Kavanagh-Baird, and B. Eskenazi. 2013. Variability of organophosphorous pesticide metabolite levels in spot and 24-hr urine samples collected from young children during 1 week. Environ. Health Perspect. 121(1):118-124.

Bruner-Tran, K.L., J. Gnecco, T. Ding, D.R. Glore, V. Pensabene, and K.G. Osteen. 2017. Exposure to the environmental endocrine disruptor TCDD and human reproductive dysfunction: Translating lessons from murine models. Reprod. Toxicol. 68:59-71

Burney, R.O., and L.C. Giudice. 2012. Pathogenesis and pathophysiology of endometriosis. Fertil. Steril. 98(3):511-519.

CDC (Centers for Disease Control and Prevention). 2009. Fourth National Report on Human Exposure to Environmental Chemicals. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA [online]. Available: https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf/ [accessed July 21, 2016].

CDC. 2015. Fourth National Report on Human Exposure to Environmental Chemicals (Updated Tables) [online]. Available: http://www.cdc.gov/biomonitoring/pdf/FourthReport_UpdatedTables_Feb2015.pdf [accessed July 1, 2016].

Connolly, R.B., G.T. Ankley, W.Y. Cheng, M.L. Mayo, D.H. Miller, E.J. Perkins, D.L. Villeneuve, and K.H. Watanabe. 2017. Quantitative adverse outcome pathways and their application to predictive toxicology. Environ. Sci. Toxicol. 51:4661-4672.

Crump, K.S., C. Van Landingham, and R.O. McClellan. 2016. Influence of alternative exposure estimates in the diesel exhaust miners study: Diesel exhaust and lung cancer. Risk Anal. 36(9):1803-1812.

Diamanti-Kandarakis, E., J.P. Bourguignon, L.C. Giudice, R. Hauser, G.S. Prins, M. Soto, R.T. Zoeller, and A.C. Gore. 2009. Endocrine-disrupting chemicals: An Endocrine Society scientific statement. Endocr. Rev. 30(4):293-342.

Donald, C.E., R.P. Scott, K.L. Blaustein, M.L. Halbleib, M. Sarr, P.C. Jepson, and K.A. Anderson. 2016. Silicone wristbands detect individuals' pesticide exposures in West Africa. R. Soc. Open Sci. 3(8):160433.

Duke, J.D., X. Han, Z. Wang, A. Subhadarshini, S.D. Karnik, X. Li, S.D. Hall, Y. Jin, J.T. Callaghan, M.J. Overhage, D.A. Flockhart, R.M. Strother, S.K. Quinney, and L. Li. 2012. Literature based drug interaction prediction with clinical assessment using electronic medical records: Novel myopathy associated drug interactions. PLoS Comput. Biol. 8(8):e1002614.

EPA (U.S. Environmental Protection Agency). 2000. CatReg Software Documentation. EPA/600/R-98/053F. Office of Research and Development, National Center for Environmental Assessment, EPA, Research Triangle Park, NC.

Eskenazi, B., P. Mocarelli, M. Warner, S. Samuels, P. Vercellini, D. Olive, L.L. Needham, D.G. Patterson Jr., P. Brambilla, N. Gavoni, S. Casalini, S. Panazza, W. Turner, and P.M. Gerthoux. 2002. Serum dioxin concentrations and endometriosis: A cohort study in Seveso, Italy. Environ. Health Perspect. 110(7):629-634.

Foster, P.M. 2014. Regulatory Forum opinion piece: New testing paradigms for reproductive and developmental toxicity—the NTP modified one generation study and OECD 443. Toxicol. Pathol. 42(8):1165-1167.

Friedman, G.D., S.F. Jiang, N. Udaltsova, J. Chan, C.P. Quesenberry, Jr., and L.A. Habel. 2009. Pharmaceuticals that cause mammary gland tumors in animals: Findings in women. Brest Cancer Res. Treat. 116(1):187-194.

Gagne, J.J., X. Han, S. Hennessy, C.E. Leonard, E.A. Chrischilles, R.M. Carnahan, S.V. Wang, C. Fuller, A. Iyer, H. Katcoff, T.S. Woodworth, P. Archdeacon, T.E. Meyer, S. Schneeweiss, and S. Toh. 2016. Successful comparison of U.S. Food and Drug Administration Sentinel analysis tools to traditional approaches in quantifying a known drug-adverse event association. Clin. Pharmacol. Ther. 100(5):558-564.

Geyh, A.S., P.T. Roberts, F.W. Lurmann, B.M. Schoell, and E.L. Avol. 1999. Initial field evaluation of the Harvard active ozone sampler for personal ozone monitoring. J. Expo. Anal. Environ. Epidemiol. 9(2):143-149.

Goodman, S.N. 2016. Aligning statistical and scientific reasoning. Science 352(6290):1180-1181.

Gore, A.C., V.A. Chappell, S.E. Fenton, J.A. Flaws, A. Nadal, G.S. Prins, J. Toppari, and R.T. Zoeller. 2015. EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. Endocr. Rev. 36(6):E1-E150.

Gurulingappa, H., A. Rajput, and L. Toldo. 2012. Extraction of adverse drug effects from medical case reports. J. Biomed. Semantics 3(1):15.

Hammel, S.C., K. Hoffman, T.F. Webster, K.A. Anderson, and H.M. Stapleton. 2016. Measuring personal exposure to organophosphate flame retardants using silicone wristbands and hand wipes. Environ. Sci. Technol. 50(8):4483-4491.

Harpaz, R., W. DuMouchel, N.H. Shah, D. Madigan, P. Ryan, and C. Friedman. 2012. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin. Pharmacol. Ther. 91(6):1010-1021.

Hartle, J.C., A. Navas-Acien, and R.S. Lawrence. 2016. The consumption of canned food and beverages and urinary Bisphenol A concentrations in NHANES 2003-2008. Environ. Res. 150:375-382.

Hass, U., S. Christiansen, M. Axelstad, K.D. Sorensen, and J. Boberg. 2013. Input for the REACH-review in 2013 on Endocrine Disruptors. Danish Centre on Endocrine Disrupters [online]. Available: http://mst.dk/media/mst/9106721/rapport_input_for_the_reach-review.pdf [accessed November 3, 2016].

Hauben, M., S. Horn, and L. Reich. 2007. Potential use of data-mining algorithms for the detection of 'surprise' adverse drug reactions. Drug Saf. 30(2):143-155.

Heindel, J.J. 1988. The National Toxicology Program chemical nomination selection and testing process. Reprod. Toxicol. 2(3-4):273-279.

Hoover, R.N., M. Hyer, R.M. Pfeiffer, E. Adam, B. Bond, A.L. Cheville, T. Colton, P. Hartge, E.E. Hatch, A.L. Herbst, B.Y. Karlan, R. Kaufman, K.L. Noller, J.R. Palmer, S.J. Robboy, R.C. Saal, W. Strohsnitter, L. Titus-Ernstoff, and R. Troisi. 2011. Adverse health outcomes in women exposed in utero to diethylstilbestrol. N. Engl. J. Med. 365(14):1304-1314.

Ibrahim, H., A. Saad, A. Abdo, and A. Sharaf Eldin. 2016. Mining association patterns of drug-interactions using post marketing FDA's spontaneous reporting data. J. Biomed. Inform 60:294-308.ICASR (International Collaboration for the Automation of System Reviews). 2015. The Vienna Principles: General Principles of Collaboration and Development in the Automation of Systematic Reviews. The First ICASR Meeting, October 2, 2015, Vienna, Austria [online]. Available: http://ebrnetwork.org/wp-content/uploads/2015/04/The-Vienna-Principles.pdf [accessed March 6, 2017].

IOM (Institute of Medicine). 2012. Breast Cancer and the Environment: A Life Course Approach. Washington, DC: The National Academies Press.

Jefferson, W.N., D.M. Chevalier, J.Y. Phelps, A.M. Cantor, E. Padilla-Banks, R.R. Newbold, T.K. Archer, H.K. Kinyamu, and C.J. Williams. 2013. Persistently altered epigenetic marks in the mouse uterus after neonatal estrogen exposure. Mol. Endocrinol. 27(10):1666-1677.

Jenkins, S., C. Rowell, J. Wang, and C.A. Lamartiniere. 2007. Prenatal TCDD exposure predisposes for mammary cancer in rats. Reprod. Toxicol. 23(3):391-396.

Kim, K.N., Y.H. Choi, Y.H. Lim, and Y.C. Hong. 2016. Urinary phthalate metabolites and depression in an elderly population: National Health and Nutrition Examination Survey 2005-2012. Environ. Res. 145:61-67.

Kim, T.S., K.K. Jung, S.S. Kim, I.H. Kang, J.H. Baek, H.S. Nam, S.K. Hong, B.M. Lee, J.T. Hong, K.W. Oh, H.S. Kim, S.Y. Han, and T.S. Kang. 2010. Effects of in utero exposure to di(n-butyl)phthalate on development of male reproductive tracts in Sprague-Dawley rats. J. Toxicol. Environ. Health A 73(21-22):1544-1559.

King, C.M., C. Barbara, A. Prentice, J.D. Brenton, and D.S. Charnock-Jones. 2016. Models of endometriosis and their utility in studying progression to ovarian clear cell carcinoma. J. Pathol. 238(2):185-196.

Lardon, J., R. Abdellaoui, F. Bellet, H. Asfari, J. Souvignet, N. Texier, M.C. Jaulent, M.N. Beyens, A. Burgun, and C. Bousquet. 2015. Adverse drug reaction identification and extraction in social media: A scoping review. J. Med. Internet Res. 17(7):e171.

Lewis, R.C., L.E. Johns, and J.D. Meeker. 2015. Serum biomarkers of exposure to perfluoroalkyl substances in relation to serum testosterone and measures of thyroid function among adults and adolescents from NHANES 2011-2012. Int. J. Environ. Res. Public Health 12(6):6098-6114.

Li, Y., H. Salmasian, S. Vilar, H. Chase, C. Friedman, and Y. Wei. 2014. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. J. Am. Med. Inform. Assoc. 21(2):308-314.

Liu, C., X. Xu, and X. Huo. 2014. Anogenital distance and its application in environmental health research. Environ. Sci. Pollut. Res. Int. 21(8):5457-5464.

Macon, M.B., and S.E. Fenton. 2013. Endocrine disruptors and the breast: Early life effects and later life disease. J. Mammary Gland Biol. Neoplasia 18(1):43-61.

Makris, S.L. 2011. Current assessment of the effects of environmental chemicals on the mammary gland in guideline rodent studies by the U.S. Environmental Protection Agency (U.S. EPA), Organisation for Economic Co-operation and Development (OECD), and National Toxicology Program (NTP). Environ. Health Perspect. 119(8):1047-1052.

Marty, M.S., B. Allen, R.E. Chapin, R. Cooper, G.P. Daston, J.A. Flaws, P.N. Foster, S.L. Makris, E. Mylchreest, D. Sandler, and R.W. Tyl. 2009. Inter-laboratory control data for reproductive endpoints required in the OPPTS 870.3800/OECD 416 reproduction and fertility test. Birth Defects Res. B Dev. Reprod. Toxicol. 86(6):470-489.

McMahon, A.W., G.T. Wharton, R. Bonnel, M. DeCelle, K. Swank, D. Testoni, J.U. Cope, P.B. Smith, E. Wu, and M.D. Murphy. 2015. Pediatric post-marketing safety systems in North America: Assessment of the current status. Pharmacoepidemiol. Drug Saf. 24(8):785-792.

Mihaylov, N.L., and D.D. Perkins. 2015. Local environmental grassroots activism: Contributions from environmental psychology, sociology and politics. Behav. Sci. 5(1):121-153.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2015. Application of Modern Toxicology for Predicting Acute Toxicity for Chemical Defense. Washington, DC: The National Academies Press.

NASEM. 2017. Using 21st Century Science to Improve Risk-Related Evaluations. Washington, DC: The National Academies Press.

Nayyar, T., K.L. Bruner-Tran, D. Piestrzeniewicz-Ulanska, and K.G. Osteen. 2007. Developmental exposure of mice to TCDD elicits a similar uterine phenotype in adult animals as observed in women with endometriosis. Reprod Toxicol. 23(3):326-336.

Newbold, R.R. 2004. Lessons learned from perinatal exposure to diethylstilbestrol. Toxicol. Appl. Pharmacol. 199(2):142-150.

Newbold, R.R., and J.A. McLachlan. 1982. Vaginal adenosis and adenocarcinoma in mice exposed prenatally or neonatally to diethylstilbestrol. Cancer Res 42(5):2003-2011.

Newbold, R.R., W.N. Jefferson, E. Padilla-Banks, and J. Haseman. 2004. Developmental exposure to diethylstilbestrol (DES) alters uterine response to estrogens in prepubescent mice: Low versus high dose effects. Reprod. Toxicol. 18(3):399-406.

Nikfarjam, A., A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez. 2015. Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J. Am. Med. Inform. Assoc. 22(3):671-681.

NRC (National Research Council). 1991. Human Exposure Assessment for Airborne Pollutants: Advances and Opportunities. Washington, DC: National Academy Press.

NRC. 2006a. Human Biomonitoring for Environmental Chemicals. Washington, DC: The National Academies Press.

NRC. 2006b. Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. Washington, DC: The National Academies Press.

NRC. 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2009. Science and Decisions: Advancing Risk Assessment. Washington DC: The National Academies Press.

NRC. 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: The National Academies Press.

NRC. 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

O'Connell, S.G., L.D. Kincl, and K.A. Anderson. 2014. Silicone wristbands as personal passive samplers. Environ. Sci. Technol. 48(6):3327-3335.

Oliveira, M., K. Slezakova, C. Delerue-Matos, M. do Carmo Pereira, and S. Morais. 2016. Assessment of polycyclic aromatic hydrocarbons in indoor and outdoor air of preschool environments (3-5 years old children). Environ. Pollut. 208 (Part B):382-394.

Omoike, O.E., R.C. Lewis, and J.D. Meeker. 2015. Association between urinary biomarkers of exposure to organophosphate insecticides and serum reproductive hormones in men from NHANES 1999-2002. Reprod. Toxicol. 53:99-104.

Patel, C.J., J. Bhattacharya, and A.J. Butte. 2010. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS ONE 5(5):e10746.

Patel, C.J., T. Yang, Z. Hu, Q. Wen, J. Sung, Y.Y. El-Sayed, H. Cohen, J. Gould, D.K. Stevenson, G.M. Shaw, X.B. Ling, and A.J. Butte. 2014. Investigation of maternal environmental exposures in association with self-reported preterm birth. Reprod. Toxicol. 45:1-7.

Perera, F.P., V. Rauh, W.Y. Tsai, P. Kinney, D. Camann, D. Barr, T. Bernet, R. Garfinkel, Y.H. Tu, D. Diaz, J. Dietrich, and R.M. Whyatt. 2003. Effects of transplacental exposure to environmental pollutants on birth outcomes in a multi-ethnic population. Environ. Health Perspect.. 111(2):201-205.

Peretz, J., L. Vrooman, W.A. Ricke, P.A. Hunt, S. Ehrlich, R. Hauser, V. Padmanabhan, H.S. Taylor, S.H. Swan, C.A. VandeVoort, and J.A. Flaws. 2014. Bisphenol A and reproductive health: Update of experimental and human evidence, 2007-2013. Environ. Health Perspect. 122(8):775-786.

Platt, R., R.M. Carnahan, J.S. Brown, E. Chrischilles, L.H. Curtis, S. Hennessy, J.C. Nelson, J.A. Racoosin, M. Robb, S. Schneeweiss, S. Toh, and M.G. Weiner. 2012. The U.S. Food and Drug Administration's Mini-Sentinel program: Status and direction. Pharmacoepidemiol. Drug Saf. 21 (Suppl. 1):1-8.

Pontes, H., M. Clement, and V. Rollason. 2014. Safety signal detection: The relevance of literature review. Drug Saf. 37(7):471-479.

Puchalski, M.A., C.M. Rogers, R. Baumgardner, K.P. Mishoe, G. Price, M.J. Smith, N. Watkins, and C.M. Lehmann. 2015. A statistical comparison of active and passive ammonia measurements collected at Clean Air Status and Trends Network (CASTNET) sites. Environ. Sci.Process. Impacts 17(2):358-369.

Reed, C.E., and S.E. Fenton. 2013. Exposure to diethylstilbestrol during sensitive life stages: A legacy of heritable health effects. Birth Defects Res. C Embryo Today 99(2):134-146.

Revesz, C., C.I. Chappel, and R. Gaudry. 1960. Masculinization of female fetuses in the rat by progestational compounds. Endocrinology 66(1):140-144.

Rier, S.E., D.C. Martin, R.E. Bowman, W.P. Dmowski, and J.L. Becker. 1993. Endometriosis in rhesus monkeys (*Macaca mulatta*) following chronic exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Fundam Appl Toxicol. 21(4):433-441.

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 science assessments. Environ. Health Perspect. 122(7):711-718.

Rudel, R.A., K.R. Attfield, J.N. Schifano, and J.G. Brody. 2007. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. Cancer 109(12 Suppl.):2635-2666.

Rudel, R.A., S.E. Fenton, J.M. Ackerman, S.Y. Euling, and S.L. Makris. 2011. Environmental exposures and mammary gland development: State of the science, public health implications, and research recommendations. Environ. Health Perspect. 119(8):1053-1061.

Rudel, R.A., J.M. Ackerman, K.R. Attfield, and J.G. Brody. 2014. New exposure biomarkers as tools for breast cancer epidemiology, biomonitoring, and prevention: A systematic approach based on animal evidence. Environ. Health Perspect. 122(9):881-895.

Saillenfait, A.M., J.P. Sabaté, and F. Gallissot. 2008. Diisobutyl phthalate impairs the androgen-dependent reproductive development of the male rat. Reprod Toxicol. 26(2):107-115.

Salazar-Martinez, E., P. Romano-Riquer, E. Yanez-Marquez, M.P. Longnecker, and M. Hernandez-Avila. 2004. Anogenital distance in human male and female newborns: A descriptive, cross-sectional study. Environ. Health 3(1):8.

Sarkar, U., A. Lopez, J.H. Maselli, and R. Gonzalez. 2011. Adverse drug events in U.S. adult ambulatory medical care. Health. Serv. Res 46(5):1517-1533.

Sarker, A., and G. Gonzalez. 2015. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J. Biomed. Inform. 53:196-207.

Sather, M.E., and K. Cavender. 2016. Trends analyses of 30 years of ambient 8 hour ozone and precursor monitoring data in the South Central U.S.: Progress and challenges. Environ. Sci. Process.Impacts 18(7):819-831.

Sathyanarayana, S., R. Grady, J.B. Redmon, K. Ivicek, E. Barrett, S. Janssen, R. Nguyen, and S.H. Swan. 2015. Anogenital distance and penile width measurements in The Infant Development and the Environment Study (TIDES): Methods and predictors. J. Pediatr. Urol. 11(2):76.e1-6.

Scholze, M., and A. Kortenkamp. 2007. Statistical power considerations show the endocrine disruptor low-dose issue in a new light. Environ. Health Perspect. 115(Suppl. 1):84-90.

Schotland, P., N. Bojunga, A. Zien, M.N. Trame, and L.J. Lesko. 2016. Improving drug safety with a systems pharmacology approach. Eur. J. Pharm. Sci. 94:84-92.

Schwarzman, M.R., J.M. Ackerman, S.H. Dairkee, S.E. Fenton, D. Johnson, K.M. Navarro, G. Osborne, R.A. Rudel, G.M. Solomon, L. Zeise, and S. Janssen. 2015. Screening for chemical contributions to breast cancer risk: A case study for chemical safety evaluation. Environ. Health Perspect. 123(12):1255-1264.

Shetty, K.D., and S.R. Dalal. 2011. Using information mining of the medical literature to improve drug safety. J. Am. Med. Inform. Assoc. 18(5):668-674.

Skakkebaek, N.E., J. Toppari, O. Söder, C.M. Gordon, S. Divall, and M. Draznin. 2011. The exposure of fetuses and children to endocrine disrupting chemicals: A European Society for Paediatric Endocrinology (ESPE) and Pediatric Endocrine Society (PES) call to action statement. J. Clin. Endocrinol. Metab. 96(10):3056-3058.

Slayden, O.D. 2016. Translational in vivo models for women's health: The nonhuman primate endometrium—A predictive model for assessing steroid receptor modulators. Handb. Exp. Pharmacol. 232:191-202.

Smith, M.T., K.Z. Guyton, C.F. Gibbons, J.M. Fritz, C.J. Portier, I. Rusyn, D.M. DeMarini, J.C. Caldwell, R.J. Kavlock, P.F. Lambert, S.S. Hecht, J.R. Bucher, B.W. Stewart, R.A. Baan, V.J. Cogliano, and K. Straif. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ. Health Perspect. 124(6):713-721.

Sobus, J.R., R.S. DeWoskin, Y.M. Tan, J.D. Pleil, M.B. Phillips, B.J. George, K. Christensen, D.M. Schreinemachers, M.A. Williams, E.A. Cohen Hubal, and S.W. Edwards. 2015. Uses of NHANES biomarker data for chemical risk assessment: Trends, challenges, and opportunities. Environ. Health Perspect. 123(10):919-927.

Solomon, P.A., and C. Sioutas. 2008. Continuous and semicontinuous monitoring techniques for particulate matter mass and chemical components: A synthesis of findings from EPA's Particulate Matter Supersites Program and related studies. J. Air Waste Manage. Assoc.58(2):164-195.

Subramaniam, R.P., K.S. Crump, C. Van Landingham, P. Whiate, C. Cehn, and P.M. Schlosser. 2007. Uncertainties in the CIIT model for formaldehyde-induced carcinogenicity in the rat: A limited sensitivity analysis-I. Risk Anal. 27(5):1237-1254.

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, and J.L. Teague. 2005. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ. Health Perspect. 113(8):1056-1061.

Tan, Y.M., J. Sobus, D. Chang, R. Tornero-Velez, M. Goldsmith, J. Pleil, and C. Dary. 2012. Reconstructing human exposures using biomarkers and other "clues." J. Toxicol. Environ. Health B Crit. Rev.15(1):22-38.

Tatonetti, N.P., G.H. Fernald, and R.B. Altman. 2012. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. J. Am. Med. Inform. Assoc. 19(1):79-85.

Taylor, R.N., M.A. Kane, and N. Sidell. 2015. Pathogenesis of endometriosis: Roles of retinoids and inflammatory pathways. Semin. Reprod. Med. 33(4):246-256.

Teeguarden, J., S. Hanson-Drury, J.W. Fisher, and D.R. Doerge. 2013. Are typical human serum BPA concentrations measurable and sufficient to be estrogenic in the general population? Food. Chem. Toxicol. 62:949-963.

Teeguarden, J.G., N.C. Twaddle, M.I. Churchwell, X. Yang, J.W. Fisher, L.M. Seryak, and D.R. Doerge. 2015. 24-hour human urine and serum profiles of bisphenol A: Evidence against sublingual absorption following ingestion in soup. Toxicol. Appl. Pharmacol. 288(2):131-142.

Thayer, K.A., and P.M. Foster. 2007. Workgroup report: National Toxicology Program Workshop on Hormonally Induced Reproductive Tumors—Relevance of Rodent Bioassays. Environ. Health Perspect.115(9):1351-1356.

Thayer, K.A., D.R. Doerge, D. Hunt, S.H. Schurman, N.C. Twaddle, M.I. Churchwell, S. Garantziotis, G.E. Kissling, M.R. Easterling, J.R. Bucher, and L.S. Birnbaum. 2015. Pharmacokinetics of bisphenol A in humans following a single oral administration. Environ. Int. 83:107-115.

Trifirò, G., A. Pariente, P.M. Coloma, J.A. Kors, G. Polimeni, G. Miremont-Salamé, M.A. Catania, F. Salvo, A. David, N. Moore, A.P. Caputi, M. Sturkenboom, M. Molokhia, J. Hippisley-Cox, C.D. Acedo, J. van der Lei, and A. Fourrier-Reglat. EU-ADR group. 2009. Data mining on electronic health record databases for signal detection in pharmacovigilance: Which events to monitor? Pharmacoepidemiol. Drug Saf. 18(12):1176-1184.

Tsai, D.H., N. Amyai, P. Marques-Vidal, J.L. Wang, M. Riediker, V. Mooser, F. Paccaud, G. Waeber, P. Vollenweider and M. Bochud. 2012. Effects of particulate matter on inflammatory markers in the general adult population. Part Fibre Toxicol. 9:24.

Vandenberg, L.N., S. Ehrlich, S.M. Belcher, N. Ben-Jonathan, D.C. Dolinoy, E.R. Hugo, P.A. Hunt, R.R. Newbold, B.S. Rubin, K.S. Saili, A.M. Soto, H.-S. Wang, and F.S. vom Saal. 2013. Low dose effects of bisphenol A. Endocr. Disruptors 1(1):e26490.

Vercellini, P., P. Viganò, E. Somigliana, and L. Fedele. 2014. Endometriosis: Pathogenesis and treatment. Nat. Rev. Endocrinol. 10(5):261-275.

Wang, W., K. Haerian, H. Salmasian, R. Harpaz, H. Chase, and C. Friedman. 2011. A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations. AMIA Annu. Symp. Proc. 2011:1464-1470.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

WHO (World Health Organization). 2002. The Importance of Pharmacovigilance—Safety Monitoring of Medicinal Products [online]. Available: http://apps.who.int/medicinedocs/en/d/Js4893e/1.html [accessed July 21, 2016].

Winnenburg, R., A. Sorbello, A. Ripple, R. Harpaz, J. Tonning, A. Szarfman, H. Francis, and O. Bodenreider. 2015. Leveraging MEDLINE indexing for pharmacovigilance—Inherent limitations and mitigation strategies. J. Biomed. Inform. 57:425-435.

Woodruff, T.J., and P. Sutton. 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(10):1007-1014.

Yang, X., D.R. Doerge, J.G. Teeguarden, and J.W. Fisher. 2015. Development of a physiologically based pharmacokinetic model for assessment of human exposure to bisphenol A. Toxicol. Appl. Pharmacol. 289(3):442-456.

Yoon, M., J.L. Campbell, M.E. Andersen, and H.J. Clewell. 2012. Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. Crit. Rev. Toxicol. 42(8):633-652.

Zoller, H.M. 2017. Health activism targeting corporations: A critical health communication perspective. Health Commun. 32(2):219-229.

3

Phthalates and Male Reproductive-Tract Development

Phthalates are ubiquitous environmental chemicals that are anti-androgenic. They are found in a wide variety of consumer products, including toys, cosmetics, pharmaceuticals, and building and construction materials. Human exposure to phthalates has been well documented and occurs following ingestion, dermal exposure, or inhalation (Hauser and Calafat 2005; Lioy et al. 2015). Because of concerns about the toxicity of phthalates, the use of certain phthalates in children's toys and child care articles has been regulated in the United States.[1] The European Union has also regulated the use of certain phthalates in toys, food-packaging materials, and cosmetics (EU 2004, 2005a,b, 2007). The National Health and Nutrition Examination Survey (NHANES) has documented widespread exposure to multiple phthalates in the general population (CDC 2009, 2015). An examination of temporal trends in phthalate exposure between 2001 and 2010 found reductions in the concentrations of some urinary phthalate metabolites and increases in the metabolite concentrations of replacement phthalates (Zota et al. 2014). Phthalates cross the placenta (Saillenfait et al. 1998; Fennell et al. 2004), and multiple phthalates have been measured in human and animal amniotic fluid (Silva et al. 2004; Calafat et al. 2006; Wittassek et al. 2009; Huang et al. 2016). In the rat, alterations in male reproductive-tract development after in utero exposure are the most sensitive health outcomes resulting from exposure to phthalates (NRC 2008; CHAP 2014). In rats, the anti-androgenic phthalates are those with ester side chains containing 4-10 carbon atoms, and some phthalates (e.g., dimethyl and diethyl phthalate) are not anti-androgenic or reproductive toxicants in the male rat (Gray et al. 2000; Furr et al. 2014).

Diester phthalates are initially hydrolyzed to their monoester, which undergoes subsequent glucuronidation and urinary excretion (see Figure 3-1). Other phthalate monoester metabolites can undergo additional oxidation of the alkyl side chains resulting in more complex metabolic profiles (Latini 2005; Calafat et al. 2006). For example, di(2-ethylhexyl)phthalate (DEHP) is metabolized to mono-2-ethylhexyl phthalate (MEHP), which undergoes additional oxidative side chain metabolism. Some representative phthalates and their oxidative metabolites are provided in Table 3-1. Biomonitoring efforts rely on the measurement of metabolite concentrations in urine (Samandar et al. 2009; Johns et al. 2015).

Male reproductive outcomes found in animal studies from in utero exposure to phthalates has been referred to as "phthalate syndrome" and include decreased anogenital distance (AGD), infertility, decreased sperm count, cryptorchidism (undescended testes), hypospadias (malformation of the penis in which the urethra does not open at the tip of the organ), and other reproductive-tract malformations (Gray et al. 2000; Foster et al. 2001; Fisher et al. 2003; NRC 2008). A hypothesized syndrome in the human ("testicular dysgenesis syndrome") shares some of the same end points as the rat phthalate syndrome (Skakkebaek 2002; NRC 2008; Wohlfahrt-Veje et al. 2009); the etiology of the proposed human syndrome is unknown, however, and may or may not involve exposure to phthalates.

Phthalate male reproductive toxicity was one of the case examples the committee explored at a workshop held on February 3, 2016, which was designed to assist the committee with selecting the topics for its systematic reviews (see Appendix B for the workshop agenda and topics). Positive feedback was received from the participants at the meeting that there is an adequate data set to perform systematic reviews of the animal and the human evidence and to explore dose-response relationships on the effects of phthalates on male reproductive-tract development.

[1]Consumer Product Safety Improvement Act of 2008, Title II § 108 (a)(b) (H.R. 4040).

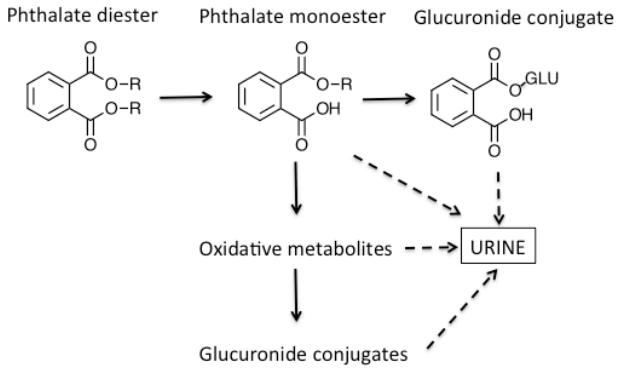


FIGURE 3-1 Overview of phthalate metabolism in mammals. Source: Adapted from Li et al. (2014).

TABLE 3-1 Parent Phthalate and Oxidative Metabolites Found in Urine Following Exposure

Phthalate (CAS no.; MW)	Abbreviation	Urinary Metabolite (CAS no.)	Abbreviation
Dimethyl phthalate (131-11-3; 194.2)	DMP	Mono-methyl phthalate (4376-18-5)	MMP
Diethyl phthalate (84-66-2; 222.2)	DEP	Mono-ethyl phthalate (2306-33-4)	MEP
Dibutyl phthalate (84-74-2; 278.3)	DBP	Mono-*n*-butyl phthalate (131-70-4)	MBP
		Mono-isobutyl phthalate (30833-53-5)	MIBP
Diisobutyl phthalate (84-69-5; 278.4)	DIBP	Mono-*n*-butyl phthalate (131-70-4)	MnBP
		Mono-isobutyl phthalate (30833-53-5)	MIBP
Dipentyl phthalate (131-18-0; 306.4)	DPP	Mono(4-hydroxypentyl) phthalate (1334312-05-8)	MHPP
		Mono(4-carboxybutyl) phthalate (92569-48-7)	MCBP
		Mono-*n*-pentyl phthalate (24539-56-8)	MPP
Benzylbutyl phthalate (85-68-7; 312.4)	BzBP	Mono-benzyl phthalate (2528-16-7)	MBzP
		Mono-*n*-butyl phthalate (131-70-4)	MnBP
Diethylhexyl phthalate (117-81-7; 390.6)	DEHP	Mono-2-ethylhexyl phthalate (4376-20-9)	MEHP
		Mono-(2-ethyl-5-hydroxyhexyl) phthalate (40321-99-1)	MEHHP
		Mono-(2-ethyl-5-oxohexyl) phthalate (40321-98-0)	MEOHP
		Mono-(2-ethyl-5-caroxypentyl) phthalate (40809-41-4)	MECPP
Diisooctyl phthalate (27554-26-3; 390.6)	DIOP	Mono-diisooctyl phthalate Mono-(3-carboxypropyl) phthalate (66851-46-5)	MDiOP MCPP
Di-*n*-octyl phthalate (117-84-0; 390.6)	DOP	Mono-(3-carboxypropyl) phthalate (66851-46-5)	MCPP
		Mono-*n*-octyl phthalate (5393-19-1)	MOP
Diisononyl phthalate (28553-12-0; 418.6)	DINP	Mono-isononyl phthalate	MINP
		Mono-carboxy-isooctyl phthalate (898544-09-7)	MCOP
Diisodecyl phthalate (26761-40-0; 446.7)	DIDP	Mono-(carboxynonyl) phthalate	MCNP

The committee focused its review on end points relevant to the anti-androgenic activity of phthalates, including fetal testosterone concentration, AGD, and hypospadias. A mechanistic link between decreased fetal testosterone levels and AGD and hypospadias is well established in animal models (e.g., Wilson et al. 2008; Scott et al. 2009). In rats, AGD is a well-known marker of androgen activity during the male programming window. Although the cause of hypospadias in humans can be multifactorial, mutations reducing androgen activity cause hypospadias in humans (van der Zanden et al. 2012). An association between hypospadias and reduced AGD has been observed in humans (Hsieh et al. 2012; Jain and Singal 2013; Thankamony et al. 2014), which suggests that human AGD is also dependent on androgen activity during the human male programming window—that is, the period during gestation when the male reproductive tract is programmed so that it will differentiate and grow normally (Ban et al. 2008; Hsieh et al. 2012; Dean and Sharpe 2013). The male programming window in the rat is gestation days 16-18, which corresponds to gestation days 14-16 in the mouse and approximately gestation weeks 8-14 in the human (Welsh et al. 2008).

Consideration was given to including cryptorchidism as an end point, but the committee decided against it for several reasons. Mechanisms for phthalate-induced cryptorchidism involve not only reduced fetal testis testosterone production but also reductions in fetal testis insulin-like 3 (INSL-3) production (Howdeshell et al. 2015). Rats exposed to phthalates have similar sensitivity to decreased fetal testosterone and AGD just as they do for decreased INSL-3 (Gray et al. 2016). In addition, cryptorchidism is a less sensitive end point compared to reductions in AGD (Saillenfait et al. 2008; Kim et al. 2010). Few human studies that examined the relationship between phthalate exposure and cryptorchidism were available to compare with animal data, which was also an important comparison to address the committee's statement of task. Because the committee's objective was to use the results of animal and human systematic reviews to evaluate the coherence between effects and dose-response relationships, the committee judged that including cryptorchidism in the analysis would not provide additional value to the project.

Two systematic reviews were conducted to answer the question what is the effect of in utero exposure to phthalates on AGD, hypospadias, or testosterone concentrations in males? One systematic review focused on animal studies and the other on human studies. This chapter first presents the methods that were used to conduct the two reviews. Then results of the reviews are presented together, along with mechanistic and other relevant information, to draw hazard conclusions.

SYSTEMATIC REVIEW METHODS

Protocols for the conduct of the systematic reviews were developed and peer reviewed. The PECO (Population, Exposure, Comparator, and Outcome) statements for the systematic reviews of the animal and the human studies are presented in Boxes 3-1 and 3-2, and the protocols used to conduct the systematic reviews are provided in Appendix C (Section C-1) and Appendix D (Section D-1), respectively. The protocols were based on the method developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) for conducting systematic reviews (hereto referred to as the OHAT method) (NTP 2015). A summary of the methods is briefly described below. The protocols were peer reviewed in accordance with standard report review practices of the National Academies of Sciences, Engineering, and Medicine. Most of the peer reviewers of the protocols were also peer reviewers of this report to ensure that the original protocols were followed and that any revisions or updates have been appropriately documented and justified. See the Acknowledgments for the list of peer reviewers.

Literature Searches and Screening

Scientific literature databases were searched for relevant studies on the effects of phthalates on male reproductive-tract development. A librarian, with specific training and expertise in performing searches for systematic reviews, developed and conducted the searches. A search for relevant existing systematic reviews was performed first, to avoid duplicating any recent work or work in progress. PubMed was

BOX 3-1 PECO Statement for the Phthalate (Animal) Systematic Review

Population: Nonhuman male mammals

Exposure:

- In utero exposure to any of the following ortho-phthalates or the corresponding monoester or oxidative metabolites: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS no. 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or dipentyl phthalate (CAS no. 131-18-0).
- Oral route of exposure.

Comparator: Male nonhuman mammals exposed in utero to different doses of phthalates or vehicle-only treatment.

Outcomes:

- Anogenital distance (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:
  - Anogenital index (AGI): AGD measurement divided by body weight or by the cube root of body weight.
  - Anoscrotal distance (ASD): the measured distance between the anus and base of the scrotum.
  - Anopenile distance (APD): the measured distance from the anus to the base of the penis.
- Hypospadias (incidence and severity/grade).
- Fetal testosterone concentration (e.g., measured from testes, serum, or plasma taken in utero).

BOX 3-2 PECO Statement for the Phthalate (Human) Systematic Review

Population: Male humans

Exposure:

- In utero exposure to any of the following ortho-phthalates or the corresponding monoester or oxidative metabolites: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS no. 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or dipentyl phthalate (CAS no. 131-18-0).
- No restrictions based on route of exposure. Measurements must be based on biomonitoring data (e.g., urinary monoester or oxidative metabolites, amniotic fluid oxidative phthalate metabolites, oxidative metabolites in other matricies).

Comparator: Male humans exposed in utero to lower concentrations of phthalates.

Outcomes:

- Anogenital distance (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:
  - Anogenital index (AGI): AGD measurement divided by body weight or by the cube root of body weight.
  - Anoscrotal distance (ASD): the measured distance between the anus and base of the scrotum.
  - Anopenile distance (APD): the measured distance from the anus to the base of the penis.
- Hypospadias (incidence, prevalence, and severity/grade) based on clinical guidelines for assessment.
- Testosterone concentrations measured during gestation or at delivery.

searched for systematic reviews published in 2013 or later, and the systematic-review protocol registries PROSPERO and CAMARADES were searched on August 3, 2016, for relevant protocols. Searches to support the systematic reviews were performed by the librarian in PubMed, Embase, and Toxline on August 15, 2016. The search strategies for animal and human publications are presented in the respective protocols (see Appendix C, Section C-1b, and Appendix D, Section D-1b).

References were screened at the title and abstract level and at the full-text level by the same two people using DistillerSR (https://www.evidencepartners.com). The screening criteria used are specified in the protocols (see Appendix C, Section C-1c, and Appendix D, Section D-1c). At the title and abstract screening level, if there was disagreement between the reviewers or an abstract was not available, the reference was passed on to the full-text screening level for further review. At the full-text level, disagreements about whether to include a reference were discussed by the two reviewers to reach agreement; if consensus could not be reached, a third team member was consulted to resolve the differences.

Data Extraction

Data from the included studies were entered into the Health Assessment Workspace Collaborative (HAWC), a Web-based interface application for warehousing data and creating visualizations (https://hawcproject.org). See Appendix C (Section C-1d) and Appendix D (Section D-1d) for data extraction elements for animal and human studies, respectively. One person entered data and a second person 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).

Risk of Bias and Study Quality Evaluations

Risk of bias is related to the internal validity of a study and reflects study design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect. Internal validity or risk of bias was assessed for individual studies using a tool developed for the OHAT method that outlines an approach to evaluating risk of bias for experimental animal and human studies (NTP 2015). The risk of bias criteria were customized from the basic OHAT method and described in the protocol for addressing the specific research question for this review (e.g., methods for measuring AGD and fetal testosterone) (see Appendix C, Section C-1e, and Appendix D, Section D-1e). Key risk of bias elements in animal studies included reliability of the outcome measure, blinding of researchers to treatment groups, and the issue of whether investigators controlled for litter effects in their experimental design or statistical approaches. Key risk of bias elements in human epidemiologic studies included confounding, exposure characterization, and outcome assessment (including blinding of outcome assessors). Two committee members independently assessed each study and answered all applicable risk of bias questions following prespecified criteria detailed in the study protocol. One individual from each pair then reconciled any discrepancies with input from the second committee member. Any members who were the study author of a publication under review recused themselves from the evaluation of their study.

Data Analysis and Evidence Synthesis

For each outcome, the body of evidence was synthesized qualitatively and, where appropriate, a meta-analysis was performed. If a meta-analysis was performed, summaries of main characteristics for each included study was compiled and reviewed by two team members 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 include the following:

- Experimental design (e.g., acute, chronic, multigenerational);
- Animal model used (e.g., species, strain, genetic background);
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status);
- Developmental stage of animals at treatment and outcome assessment;
- Dose levels, frequency of treatment, timing, duration, and exposure route;
- Health outcome(s) reported and their measurement;
- Type of data (e.g., continuous or dichotomous), statistics presented in the original publication; and
- Variation in degree of risk of bias at individual study level.

Uses of meta-analyses and meta-regression of experimental animal studies is provided in Box 3-3, and the methods used for performing meta-analyses, meta-regression, and benchmark dose estimation are summarized in Box 3-4.

The main characteristics considered across all eligible human studies include the following:

- Study design (e.g., cross-sectional, cohort);
- Details on how participants were classified into exposure groups (e.g., quartiles of exposure);
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring);
- Measurement of biomonitoring data specific to phthalate exposure for each exposure group;
- Health outcome(s) reported;
- Conditioning variables in the analysis (e.g., variables considered confounders);
- Type of data (e.g., continuous or dichotomous), statistics presented in paper; and
- Variation in degree of risk of bias at individual study level.

Confidence Rating and Level of Evidence Conclusions

The quality of evidence for each outcome was evaluated using a grading system based on a modification of the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). The process for rating the body of evidence as high, moderate, low, or very low was guided by the OHAT method (see Figure 3-2). In brief, studies on a particular outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Several factors were then considered to determine whether the initial rating should be downgraded or upgraded. Factors that decrease confidence in results and lead to downgrading are risk of bias, unexplained inconsistency in results, indirectness or lack of applicability, imprecision, and publication bias. Factors that increase confidence in results and can upgrade a rating are these: a large magnitude of effect; evidence of a dose-response relationship; consistency across study designs, populations, animal models, or species; consideration of residual confounding; and other factors that increase confidence in the association or effect (e.g., rare outcomes). Confidence ratings were independently assessed by two committee members, and discrepancies were resolved by consensus and consultation with a third team member as needed. After a final confidence rating is determined, the rating is translated into a level of evidence using the scheme presented in Figure 3-3.

Integration of Evidence and Drawing Hazard Identification Conclusions

The committee used guidance from OHAT 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 mechanistic information. The five possible hazard conclusions are (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 has been described as having inadequate evidence, conclusions are drawn on the basis of a single evidence base. The hazard identification scheme is presented in Figure 3-4.

RESULTS

Literature Search and Screening Results

A search for existing systematic reviews on phthalate exposure and male reproductive-tract development in animals or humans found one publication in PubMed (Kay et al. 2014), but it was a literature review article and not a systematic review. No relevant protocols for ongoing systematic reviews were found in PROSPERO or CAMARADES.

A search of electronic databases for relevant publications to address the animal systematic review PECO statement found 1,527 unique citations (see Appendix C, Section C-2). A total of 311 publications met the criteria for full-text review, and 64 of them met the inclusion criteria for data extraction. A review of the reference lists of the 64 included studies identified an additional 16 publications that were potentially relevant. Those publications underwent the same screening process as did the publications found through database searches, and six publications met the inclusion criteria for data extraction (see Figure 3-5 for an illustration of the screening process and the exclusion criteria used at the full text screening level). Thus, animal data were extracted from 70 publications (see Box 3-5).

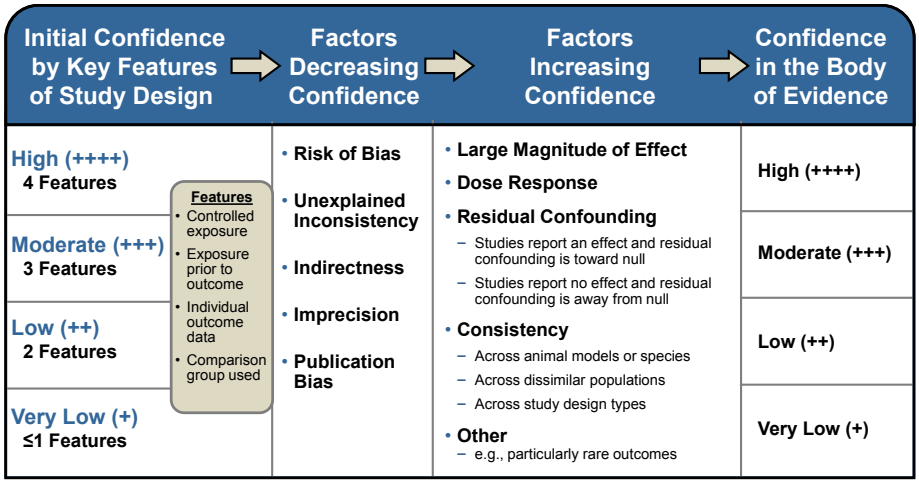
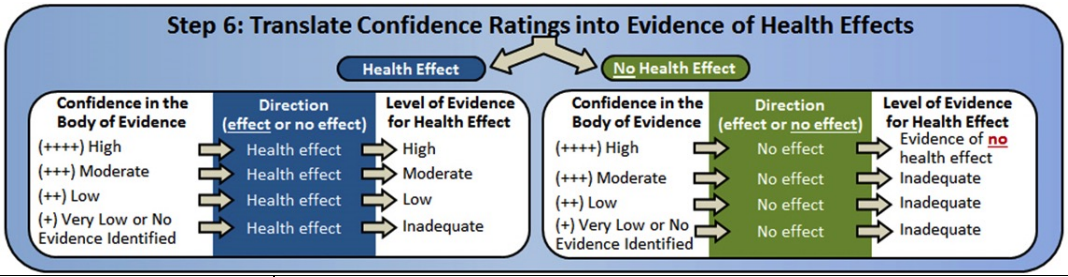


FIGURE 3-2 Method for assessing confidence in the body of evidence. SOURCE: NTP (2015).



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between exposure to the substance and the health outcome(s).
Moderate 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).
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.
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).
Inadequate Evidence	There is insufficient evidence available to assess if exposure to the substance is associated with the health outcome(s).

FIGURE 3-3 Method for translating confidence ratings into evidence for health effects. SOURCE: NTP (2015).

BOX 3-3 Uses of Meta-Analyses and Meta-Regression of Experimental Animal Studies

Background

Meta-analysis is a statistical procedure that summarizes the outcomes from a group of studies. One of the key strengths of meta-analysis is the ability to explore sources of heterogeneity through use of random effects models (NRC 2006). In that vein, ***meta-regression*** involves statistical modeling of this heterogeneity using potential explanatory variables, such as dose level or other covariates. There is a long history of performing meta-analysis for human epidemiology and clinical trial data, but its benefits for experimental animal studies are not widely recognized. Sena et al. (2014) reviewed the rationale for performing systematic review and meta-analysis of preclinical data.

Informing Confidence in the Body of Evidence for Causality

In the context of the systematic review, meta-analysis and meta-regression are useful in evaluating several factors that might increase or decrease the confidence in a body of evidence: imprecision, unexplained inconsistency in results, large magnitude of effect, and evidence of a dose-response relationship. The committee's analyses are mapped to these factors as follows:

Imprecision refers to the degree of uncertainty surrounding an effect estimate (NTP 2015). The standard practice is that the 95% confidence interval (CI) is used to assess imprecision—i.e., a downgrade for imprecision would be supported if the 95% CI overlaps with no effect (Guyatt et al. 2011). Because the hazard identification step involves evaluation of causality (i.e., whether the effect of exposure is zero or not with treatment), the meta-analysis results for the overall effect of *any* treatment, regardless of dose level, are used to assess imprecision. Additionally, the robustness of the estimate and its CI is assessed using sensitivity analyses (see Box 3-4).

Unexplained inconsistency refers to statistical heterogeneity across effect estimates that cannot be explained, for instance by random errors. As outlined in the protocols, the standard practice is to use Cochran's Q and the I^2 index to evaluate heterogeneity. If heterogeneity is observed in the overall effect, then either subgroup analyses or meta-regression based on covariates (listed in the protocol) are used to see if they can explain the observed heterogeneity. Specifically, for dose level, meta-regression was performed assuming a linear relationship with dose in original and log-transformed units, as well as a linear-quadratic relationship consisting of dose and dose-squared (in original units).

Large magnitude of effect refers to an effect estimate of a large enough size to reduce the likelihood that it could be explained by chance, confounding, or other biases. Again, because the hazard identification step involves evaluation of causality, the meta-analysis results for the overall effect of *any* treatment is used to assess whether there is a large magnitude of effect.

Evidence of a dose-response relationship refers to the presence of a dose-response gradient, which also reduces the likelihood that the effect can be explained by chance, confounding, or other biases. The purpose in this case is not to assess whether a particular dose-response shape is correct, but rather to assess whether a plausible gradient exists. Thus, the linear meta-regressions with dose in original and log-transformed units was used to assess the presence or absence of a gradient.

Characterizing Dose Response

Unlike the usual procedure of conducting dose-response analyses for each study individually, meta-regression characterizes the dose-response relationship across a group of studies, taking into account heterogeneity between and within studies through random effects. Therefore, the committee used the meta-regression results for the linear and linear-quadratic models to estimate benchmark doses that reflect the common dose-response relationship across studies, eliminating the estimated random effects of within- and between-study variation.

BOX 3-4 Summary of Meta-Analyses, Meta-Regression, and Benchmark Dose Estimation Methods for Experimental Animal Studies

Effect sizes

Meta-analysis requires that a common measure of effect size be calculated for each study and treatment group. The dichotomous data on hypospadias from phthalate experiments had too many zero count cells (incidence = 0), including all controls, which made them unsuitable for most existing methods for calculating effect sizes. The remaining continuous data had effect sizes for each treatment group calculated as follows:

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

For small differences between treated and control groups, $y_i \approx \%\ change_i$, with the exact inverse relationship being:

$$\%\ change_i = 100 \times (e^{y_i/100} - 1).$$

For instance, a –5% change corresponds to $y = -5.1$. This transformation results in confidence intervals that are more symmetric and closer to normal (Hedges et al. 1999; Lajeunesse 2011).

Meta-analysis

A standard random effects model was applied, using the Restricted Maximum Likelihood Estimate as implemented in the R package "metafor" (Viechtbauer 2005; Raudenbush 2009; Viechtbauer 2010):

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

Here 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 $u_i \sim N(0,\tau^2)$; and $\varepsilon_i \sim N(0,v_i)$ is the sampling error, where variance, v_i is calculated based on the reported sample sizes and standard deviations of the treatment and control groups. Importantly, 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 involves choosing only the single highest dose from each study, so that each y_i represents a separate study, and are 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 is estimates along with its 95% confidence interval (CI) and z-score. For heterogeneity, τ^2 is estimated, as well as the estimated Q statistic and its p-value (whether there is statistically significant heterogeneity) and the I^2 index ($I^2 = \tau^2$/overall variance).

Meta-regression and benchmark-dose estimates

Meta-regression involves adding n_i predictors $x_{i,i}$ to the random effects model in an attempt to explain the

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

The meta-regression analyses focused on the dose-response relationship. Three models were selected for illustration, as they are easily implemented with existing software meta-analysis:

Linear: $x_{1,i} = \text{dose}_i$ Log-linear: $x_{1,i} = \log_{10}(\text{dose}_i)$ Linear-quadratic: $x_{1,i} = \text{dose}_i$; $x_{2,i} = \text{dose}^2_i$

For the linear and linear-quadratic models, the "intercept" term μ was omitted to ensure that there is no effect at dose = 0. These two models were also used to estimate benchmark dose (BMD) values based on the average true effect across studies $y_{avg}(\text{dose}) = \beta_1 \times \text{dose}$ and $y_{avg}(\text{dose}) = \beta_1 \times \text{dose} + \beta_2 \times \text{dose}^2$. As with the standard BMD methodology, AIC was used to select the preferred model.

Covariates such as species and strain were assessed by sub-group analyses, which the committee felt were easier to communicate than using meta-regression. Subgroup analyses may be less practical with smaller databases. Additionally, including covariates in meta-regression could provide more quantitative insights into the contribution of those factors to heterogeneity.

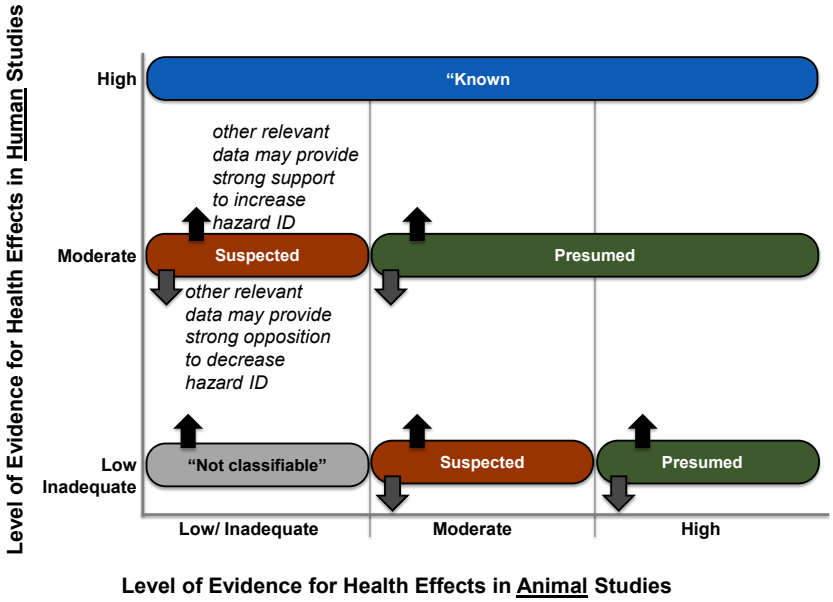


FIGURE 3-4 OHAT hazard identification scheme. SOURCE: NTP (2015).

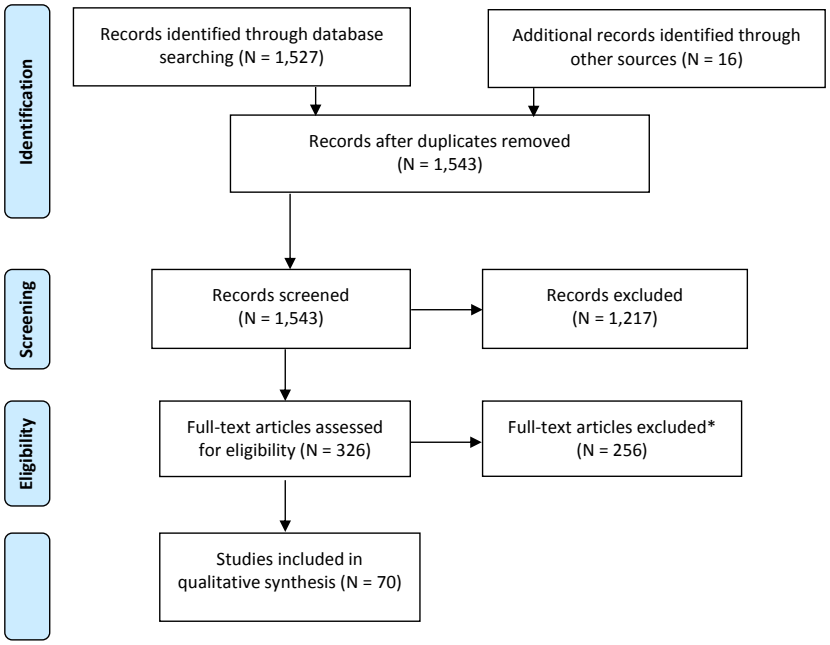


FIGURE 3-5 Summary of the search and screening of the literature on the effects of in utero exposure to phthalates on male reproductive-tract development in animals. *Articles were excluded for the following reasons: no original data (n = 8); study does not include nonhuman mammals (n = 3); study does not report experimental exposure to one or more of the phthalates listed in the PECO statement (n = 7); study does not report oral exposure to phthalates (n = 4); study does not quantify exposure to phthalates (n = 1); study does not include in utero exposure (n = 6); study does not assess or report anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or fetal testosterone concentrations (n = 73); not in English (n = 7); or other reason (n = 165). Explanations cited for exclusion because of other reasons included study involved only a single high dose (≥500 mg/kg-day), study was of exposure to a mixture, no quantitative data on anogenital measurements, no data on male animals, abstract, and duplicate. NOTE: The number of studies does not equal the total in the figure because the screeners sometimes excluded a study for different reasons.

A search of electronic databases for human studies found 594 unique citations (see Appendix D, Section D-2, for a breakdown by database). A total of 27 publications met the criteria for full-text review, and 13 of them met the inclusion criteria for data extraction. A review of the reference lists of the 13 human studies identified an additional six publications that were potentially relevant. Those publications underwent the same screening process as the publications found through database searches, and three publications met the inclusion criteria (see Figure 3-6). A closer evaluation of the set of 16 included publications revealed that three of the publications (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 publication (Swan 2008) had expanded results from an earlier cohort by Swan et al. (2005) and had a larger sample size. To avoid double-counting data from the same cohort, the reports from Adibi et al. (2015), Barrett et al. (2016), and Swan et al. (2005) were excluded from data extraction. Martino-Andrade (2016) was retained because it provided additional information beyond Swan et al. (2015) on windows of exposure during the second and third trimester. Thus, data were extracted from 13 publications (see Box 3-6).

BOX 3-5 Studies Included in the Phthalate (Animal) Systematic Review

Adamsson et al. 2009	Ema et al. 1998	Lee et al. 2004	Nagao et al. 2000
Ahmad et al. 2014	Ema et al. 2000	Lehmann et al. 2004	Pocar et al. 2012
Andrade et al. 2006	Fujii et al. 2005	Li et al. 2009	Saillenfait et al. 2008
Ashby et al. 1997	Furr et al. 2014	Li et al. 2013	Saillenfait et al. 2009
Aso et al. 2005	Giribabu et al. 2014	Li et al. 2015a	Saillenfait et al. 2011
Barlow et al. 2004	Gray et al. 2009	Li et al. 2015b	Saillenfait et al. 2013a
Beverly et al. 2014	Hannas et al. 2011a	Lin et al. 2008	Saillenfait et al. 2013b
Boberg et al. 2011	Hannas et al. 2011b	Lin et al. 2009	Scarano et al. 2010
Borch et al. 2004	Hannas et al. 2012	Liu et al. 2008	Struve et al. 2009
Borch et al. 2006	Howdeshell et al. 2008	MacLeod et al. 2010	Tyl et al. 2004
Christiansen et al. 2009	Jarfelt et al. 2005	Mahood et al. 2007	van den Driesche et al. 2012
Christiansen et al. 2010	Jiang et al. 2007	Martino-Andrade et al. 2009	Vo et al. 2009
Clewell et al. 2009	Johnson et al. 2007	Masutomi et al. 2003	Wolfe and Layton 2005
Clewell et al. 2013	Johnson et al. 2011	McKinnell et al. 2009	Wolfe and Patel 2002
Culty et al. 2008	Jones et al. 2015	Moore et al. 2001	Zhang et al. 2004
Do et al. 2012	Kim et al. 2010	Mylchreest et al. 1998	Zhang et al. 2013
Drake et al. 2009	Klinefelter et al. 2012	Mylchreest et al. 1999	
Ema and Miyawaki 2002	Kuhl et al. 2007	Mylchreest et al. 2000	

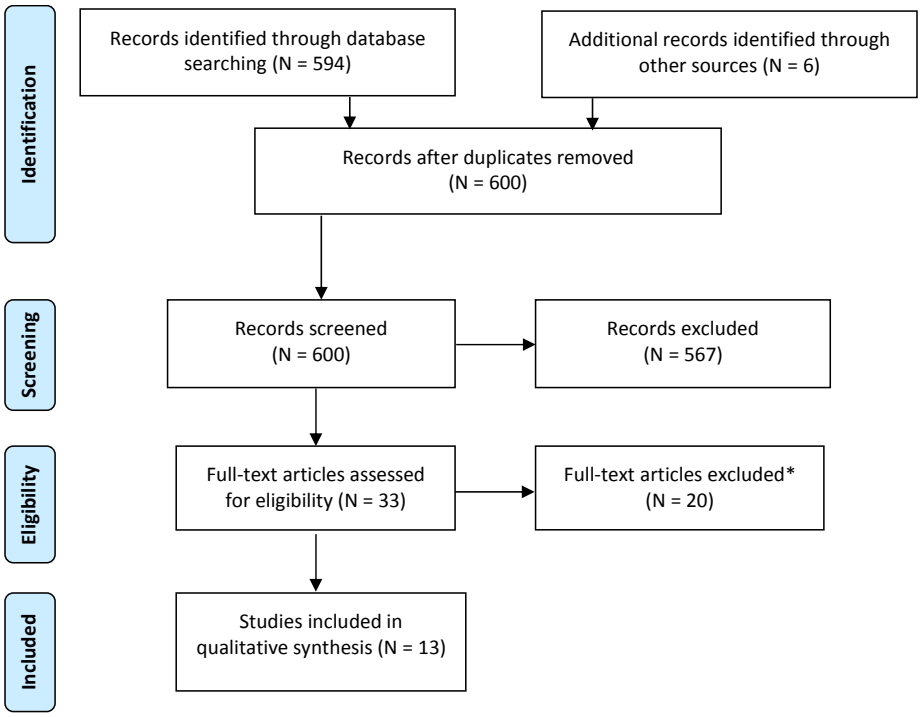


FIGURE 3-6 Summary of the search and screening of the literature on the effects of in utero exposure to phthalates on male reproductive-tract development in humans. *Articles were excluded for the following reasons: no original data (n = 7); study does not report phthalate exposure to one or more of the phthalates listed in the PECO statement (n = 2); study does not have biomonitoring data specific to phthalate exposure (n = 1); study does not assess or report anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or testosterone concentrations measured during gestation or at delivery (n = 4); or other reason (n = 10). Explanations cited for exclusion because of other reasons included no original data, duplicate, abstract, urinary measures were taken after birth, and no phthalate measure. NOTE: The number of studies does not equal the total in the figure because the screeners sometimes excluded a study for different reasons.

BOX 3-6 Studies Included in the Phthalate (Human) Systematic Review

Araki et al. 2014

Bornehag et al. 2015

Bustamante-Montes et al. 2013

Chevrier et al. 2012

Huang et al. 2009

Jensen et al. 2015

Jensen et al. 2016

Lin et al. 2011

Martino-Andrade et al. 2016

Sathyanarayana et al. 2014

Suzuki et al. 2012

Swan 2008

Swan et al. 2015

Health Effects Results

Effects of in utero phthalate exposure on male reproductive-tract development were evaluated separately for the human and the animal evidence. The outcomes examined were male AGD, fetal testosterone concentrations, and hypospadias incidence. Data were extracted from each of the studies and risk of bias assessments were performed. For the purposes of demonstrating the evaluations steps of rating the confidence in the bodies of evidence, performing qualitative and quantitative evidence syntheses, and drawing hazard identification conclusions, a decision was made to focus on a single phthalate. DEHP was selected for the example because it is a known anti-androgenic phthalate with widespread human exposures and was one of the congeners that had a robust set of human and animal studies. Results for other phthalates are also summarized later in this chapter.

Animal Health Effect Results on DEHP

Effects on AGD

Summary of the Evidence. There were 19 experimental animal studies that evaluated DEHP and AGD. Sixteen studies used the rat model and three studies used the mouse model (see Table 3-2). Phthalate exposure in all of the studies encompassed the entirety of the male programming window. Some of the data contained in the Wolfe and Layton (2005) study were from sire-only exposure and were not used in the committee's analysis. Within some studies, AGD was measured at more than one postnatal age; in these instances, only data from the earliest postnatal age were used in the analysis because AGD may change during aging (McIntyre et al. 2001). Some studies presented AGD data in more than one manner (e.g., both corrected and uncorrected for body weight); in these instances, AGD data corrected for body weight were used.

Risk of Bias Considerations. Figure 3-7 shows the risk of bias evaluations of the studies used by the committee to assess DEHP effects on AGD. The primary factors of concern for animal studies are reliability of outcome measure, blinding of researchers to treatment groups, and control for litter effects. The majority of the studies did not adequately describe 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 in most of the studies, blinding of the assessor was not reported. In addition, for the majority of studies the experimental design and/or statistical methods did not explicitly account for litter effects. Thus, most of the studies were rated as having a high risk of bias (or not reported) in these categories. The risk of bias assessment 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. Because data reporting of methods and results was often incomplete, numerous studies received a "not reported" rating for one or more of these secondary risk factors. There was no evidence of publication bias (see Appendix C, Section C-3).

Confidence in the Body of Evidence. The initial rating for the confidence in the animal studies was high because they involved controlled exposures, exposures occurred prior to outcome, outcomes were measured on individual animals, and a concurrent control comparison group was used (see Figure 3-2 for OHAT method for rating confidence). Confidence was downgraded because of the concern of significant risk of bias (described above under "Risk of Bias Considerations") related to confidence in the reliability of outcome measure, blinding of investigators to the treatment groups, and control for litter effects. Confidence was upgraded because of a large magnitude of effect and because of evidence of a dose response.

TABLE 3-2 Summary of Animal Studies of DEHP and AGD

Reference	Species/Strain	Doses, mg/kg-day	Route of Exposure	Window of In Utero Exposure	GD Definition in Study	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	Oral (gavage)	GD 6-PND 21	GD 0 = sperm positive	AGD (mm)	PND 22
Borch et al. 2004	Wistar rat	0, 750	Oral (gavage)	GD 7-21	GD 1 = day after mating	AGD (mm)	PND 3
Christiansen et al. 2009	Wistar rat	0, 3, 15, 30	Oral (gavage)	GD 7-21	GD 1 = day after mating	AGD (mm/cube root BW)	PND 0
Christiansen et al. 2010	Wistar rat: Study 1	0, 10, 30, 100, 300, 600, 900	Oral (gavage)	GD 7-PND 16	GD 1 = day after mating	AGD (mm)	PND 1
	Wistar rat: Study 2	0, 3, 10, 30, 100	Oral (gavage)	GD 7-PND 16	GD 1 = day after mating	AGD (mm)	PND 1
Culty et al. 2008	Sprague-Dawley rat: Group 1	0, 234, 469, 700, 750, 938, 1,250	Oral (gavage)	GD 14-PND 0	ND	AGD (mm)	PND 60
Do et al. 2012	CD-1 mouse	0, 0.0005, 0.001, 0.005, 0.5, 50, 500	Oral (micropipetter)	GD 9-18	GD 0 = sperm plug	AGD (mm), AGD (mm/g)	GD 18
Gray et al. 2009	Sprague-Dawley rat	0, 11, 33, 100, 300	Oral (gavage)	GD 8-PND 17	GD 1 = sperm positive	AGD (mm)	PND 2
Jarfelt et al. 2005	Wistar rat	0, 300, 750	Oral (gavage)	GD 7-PND 3	ND	AGD (mm)	PND 3
Jones et al. 2015	Sprague-Dawley rat	0, 10	Oral (gavage)	GD 14-21	ND	AGD (mm/g)	PND 3
	Sprague-Dawley rat	0, 10	Oral (gavage)	GD 14-21	ND	AGD (mm/g)	PND 6
Li et al. 2013	Sprague-Dawley rat	0, 500, 750, 1,000	Oral (gavage)	GD 12-19	GD 0 = vaginal plug	AGD (mm), AGD (mm/g)	PND 1
Lin et al. 2008	Long-Evans rat	0, 10, 100, 750	Oral (gavage)	GD 2-20	ND	AGD (mm)	GD 21
Lin et al. 2009	Long-Evans rat	0, 10, 750	Oral (gavage)	GD 12.5-PND 21	ND	AGD (mm)	PND 2
Liu et al. 2008	C57BL/6 mouse	0, 100, 200, 500	Oral (gavage)	GD 12-17	GD 0.5 = sperm positive	AGD (mm)	GD 19
Martino-Andrade et al. 2009	Wistar rat	0, 150	Oral (gavage)	GD 13-21	GD 0 = sperm positive	AGD (mm), AGD (mm/cube root BW)	GD 21
Moore et al. 2001	Sprague-Dawley rat	0, 375, 750, 1,500	Oral (ND)	GD 3-21	GD 0 = sperm positive	AGD (mm)	PND 1
Pocar et al. 2012	CD-1 mouse	0, 0.05, 5, 500	Oral (diet)	GD 0.5-PND 21	GD 0.5 = sperm positive	AGD (mm/cube root BW)	PND 42
Vo et al. 2009	Sprague-Dawley rat	0, 10, 100, 500	Oral (gavage)	GD 11-21	GD 0 = sperm positive	AGD (mm)	PND 63
Wolfe and Layton 2005	Sprague-Dawley rat (F1)	0, 321.42, 643.95	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1

	Sprague-Dawley rat (F1a-c)	0.12, 0.78, 2.37, 7.91, 23.3, 77.45, 592.3, 774.65, 0.12	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F2a-c)	0.09, 0.48, 1.4, 4.9, 14, 48, 391, 543, 0.09	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F3a-c)	0.1, 0.47, 1.4, 4.8, 14, 46, 359	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F2)—offspring of treated dams	0.09, 543	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F2)—offspring of treated sires	0.09, 543	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F3)—offspring of treated sires	0.1, 359	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
	Sprague-Dawley rat (F3)—offspring of treated dams	0.1, 359	Oral (diet)	GD 0-parturition	ND	AGD (mm), AGD (mm/g)	PND 1
Zhang et al. 2013	Sprague-Dawley rat	0, 250	Oral gavage	GD 3-PND 21	GD 0 = sperm positive	AGD (mm/cube root BW)	PND 1
	Sprague-Dawley rat	0, 250	Oral gavage	GD 3-PND 21	GD 0 = sperm positive	AGD (mm/cube root BW)	PND 22

NOTE: BW, body weight; GD, gestation day; ND, not defined; PND, postnatal day.

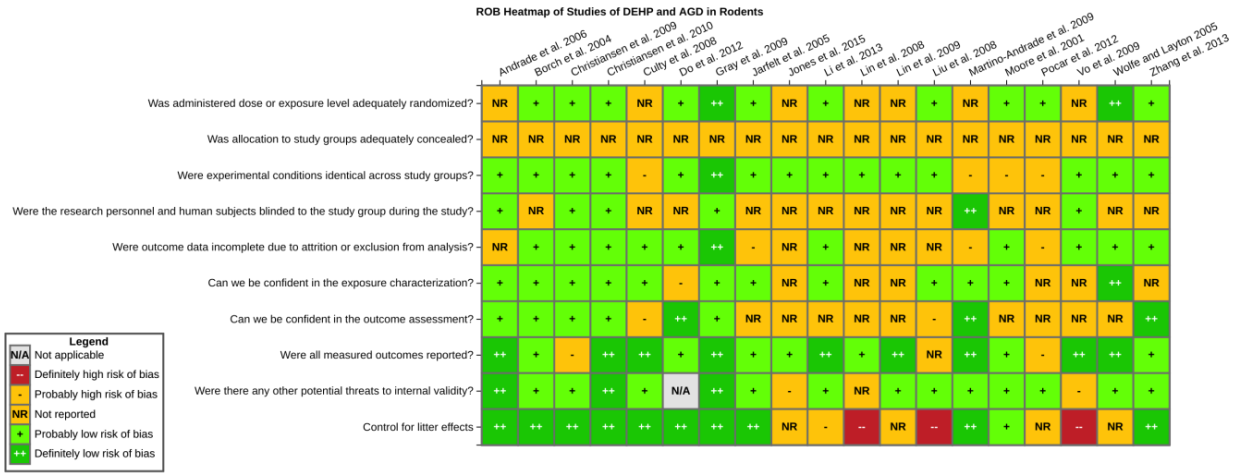


FIGURE 3-7 Risk of bias heatmap of studies of DEHP and AGD in rodents. In HAWC: https://hawcproject.org/summary/visual/361/.

A meta-analysis of studies of DEHP and AGD is presented later in this chapter (see "Meta-Analysis of DEHP and Reductions in AGD in Rats and Mice"). The results of the meta-analysis were subsequently factored into the following decisions regarding the committee's confidence in the body of evidence

- Factors potentially decreasing confidence:
  - Unexplained inconsistency. No downgrade because most of the heterogeneity can be explained by dose, species, or strain. For instance, when separated by strain, and under a linear or linear-quadratic meta-regression in dose, there is no evidence of important heterogeneity in the rat data, with low values for I^2 that were not statistically significant. Under a linear or linear-quadratic meta-regression in dose, there is no evidence of important heterogeneity in the mouse data, with I^2 values of zero.[2]
  - Imprecision. The summary overall estimate, linear trend in $\log_{10}$(dose), and linear trend in dose were all statistically significant in rats. Additionally, the statistical significance was robust under multiple sensitivity analyses. In contrast, the overall summary estimate for mice was not statistically significant; the linear trend in $\log_{10}$(dose) and in dose were both statistically significant. This statistical significance was not robust under some sensitivity analyses, however. Therefore, the meta-analysis supports a downgrade in confidence based on imprecision in the mouse studies only. Because the mouse studies account for a small percentage of the overall body of evidence (three of 19 studies) the overall confidence in the body of evidence was not downgraded for imprecision.
- Factors potentially increasing confidence
  - Large magnitude of association or effect. In rats, the effects could be considered large and robust, with overall summary estimates having z-scores[3] of ≥7.0. Moreover, these effect sizes were robust to multiple sensitivity analyses. Therefore, the meta-analysis supports an upgrade in confidence based on large magnitude of effect.

[2] I^2 describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error. The Cochrane Handbook provides the following guide to the interpretation of I^2 values: 0% to 40% (might not be important); 30% to 60% (may represent moderate heterogeneity); 50% to 90% (may represent substantial heterogeneity); and 75% to 100% (considerable heterogeneity) (Higgins and Green 2011).

[3] Z scores are calculated by the ratio of the effect estimate Beta to the standard error, which can be calculated from the 95% CI. Specifically z = Beta*3.92/(CI, upper – CI, lower). Values of Beta and the CI are given in the Appendix C, Section C-5.

- Dose response. An upgrade is supported because of strong evidence of dose response in the rat data through meta-regression with statistically significant linear trends in either $\log_{10}$(dose) or dose. Moreover, these results were robust to multiple sensitivity analyses.

Table 3-3 presents the overall high confidence rating for the body of evidence on DEHP and AGD in rodents, and the details about how the ratings was determined is presented in Appendix C, Section C-4.

Level of Evidence in the Health Effect. A meta-analysis performed on studies on AGD and DEHP (see "Meta-Analysis of DEHP and Reductions in AGD in Rats and Mice" presented later in the chapter) found consistent evidence of a decrease in AGD after in utero exposure to DEHP in rats. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence (described above) and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DEHP is associated with a reduction in AGD in male rats.

Alterations in Fetal Testosterone Concentrations

Summary of the Evidence. Twelve studies examining DEHP and fetal testosterone concentrations in animals were available (see Table 3-4), 11 in rats and one in mice. All the studies examined testosterone levels during fetal life. Unlike hypospadias and AGD analyses, studies measuring testosterone levels within fetal life but outside of the male programming window were included because fetal Leydig cell testosterone production sensitivity to phthalate exposure encompasses the entirety of fetal life when the testis is producing testosterone. The phthalate mechanism does not appear to involve an effect on pituitary-derived luteinizing hormone (Martinez-Arguelles et al. 2013); therefore, testosterone data were excluded from the analysis when the underlying fetal testis incubation method included agonism of the luteinizing hormone receptor.

TABLE 3-3 Profile of the Confidence in the Body of Evidence on DEHP and AGD in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	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
DEHP	High (16 rat,[a] 3 mouse[b])	↓	—	—	—	—	↑	↑	—	—	—	High

[a]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, 2009); Christiansen et al. (2009, 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).
[b]Liu et al. (2008); Do et al. (2012); Pocar et al. (2012).

TABLE 3-4 Summary of Animal Studies of DEHP and Testosterone

Reference	Species/Strain	Doses, mg/kg-day	Route of Exposure	Window of In Utero Exposure	GD Definition in Study	Testosterone Measurement	Age at Measurement
Borch et al. 2004	Wistar rat	0, 300, 750	Oral (gavage)	GD 7-21	GD 1 = day after mating	Testes and plasma	GD 21
	Wistar rat	0, 300, 750	Oral (gavage)	GD 7-21	GD 1 = day after mating	Testes production	GD 21 (3 h incubation)
Borch et al. 2006	Wistar rat	0, 10, 30, 100, 300	Oral (gavage)	GD 7-21	GD 1 = day after mating	Testes and plasma	GD 21
	Wistar rat	0, 10, 30, 100, 300	Oral (gavage)	GD 7-21	GD 1 = day after mating	Testes production	GD 21 (5 h incubation)
Culty et al. 2008	Sprague-Dawley rat: Group 2	0, 117, 234, 469, 938	Oral (gavage)	GD 14-20	ND	Testes production	GD 20 (24 h incubation)
Do et al. 2012	CD-1 mouse	0, 0.0005, 0.001, 0.005, 0.5, 50, 500	Oral (micropipetter)	GD 9-18	GD 0 = sperm plug	Testes and serum	GD 18
Furr et al. 2014	Sprague-Dawley rat	0, 100, 300, 600, 900	Oral (gavage)	GD 14-18	GD 0 = sperm plug	Testes production	GD 18 (3 h incubation)
Hannas et al. 2011b	Sprague-Dawley rat	0, 100, 300, 500, 625, 750, 875	Oral (gavage)	GD 14-18	GD 0/1 = sperm positive*	Testes production	GD 18 (3 h incubation)
	Wistar rat	0, 100, 300, 500, 625, 750, 875	Oral (gavage)	GD 14-18	GD 0/1 = sperm positive*	Testes production	GD 18 (3 h incubation)
Howdeshell et al. 2008	Sprague-Dawley rat	0, 100, 300, 600, 900	Oral (gavage)	GD 81-8	GD 1 = sperm plug	Testes production	GD 18 (3 h incubation)
Klinefelter et al. 2012	Sprague-Dawley rat	0, 10, 100	Oral (gavage)	GD 13-19	ND	Testes stimulated	GD 19
Lin et al. 2008	Long-Evans rat	0, 10, 100, 750	Oral (gavage)	GD 2-20	ND	Testes	GD 21
Martino-Andrade et al. 2009	Wistar rat	0, 150	Oral (gavage)	GD 13-21	GD 0 = sperm positive	Testes	GD 21
Saillenfait et al. 2013a	Sprague-Dawley rat	0, 50, 625	Oral (gavage)	GD 12-19	GD 0 = sperm positive	Testes production	GD 19 (3 h incubation)
Vo et al. 2009	Sprague-Dawley rat	0, 10, 100, 500	Oral (gavage)	GD 11-21	GD 0 = sperm positive	Serum	GD 21

*Depended on the supplier.

Risk of Bias Considerations. Figure 3-8 provides a summary of the risk of bias evaluation of the studies used by the committee to assess DEHP effects on testosterone. The primary factors of concern for animal studies are reliability of outcome measure, blinding of researchers to treatment groups, and control for litter effects. The majority of studies described the methods used to measure fetal testosterone and used measurement methods that the committee considered reliable (e.g., use of radioimmunoassay or enzyme-linked immunosorbent assay procedures). In the majority of studies, the experimental design and/or statistical methods accounted for litter effects. The risk of bias assessment also considered blinding of investigators to the treatment groups, but this factor was considered a secondary element that did not influence the committee's confidence in the body of evidence. The committee 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. Because data reporting of methods and results was often incomplete, numerous studies received a "not reported" rating for one or more of these secondary factors. There was no evidence of publication bias (see Appendix C, Section C-3).

Confidence in the Body of Evidence. The initial rating for the confidence in the animal studies was high because they involved controlled exposures, exposures occurred prior to outcome, outcomes were measured on individual animals, and a concurrent control comparison group was used (see Figure 3-2 for OHAT method for rating confidence). Confidence in the body of evidence was not downgraded for any factors, but was upgraded because of evidence of a large magnitude of effect and a dose response. A meta-analysis of studies on DEHP and fetal testosterone is presented later in this chapter (see "Meta-Analysis of DEHP and Alterations in Fetal Testosterone in Rats"), and informed decisions regarding the committee's confidence in the body of evidence, including

- Factors potentially decreasing confidence
  - Unexplained inconsistency. No downgrade was warranted because some of the heterogeneity is explained by dose, but there was substantial residual variance. The size of the effect is large enough, however, so that concerns about inconsistency were not serious from the point of view of causal inference.
  - Imprecision. No downgrade was warranted because the overall summary estimate, linear trend in $\log_{10}$(dose), and linear trend in dose were all statistically significant. Additionally, the statistical significance was robust under multiple sensitivity analyses.

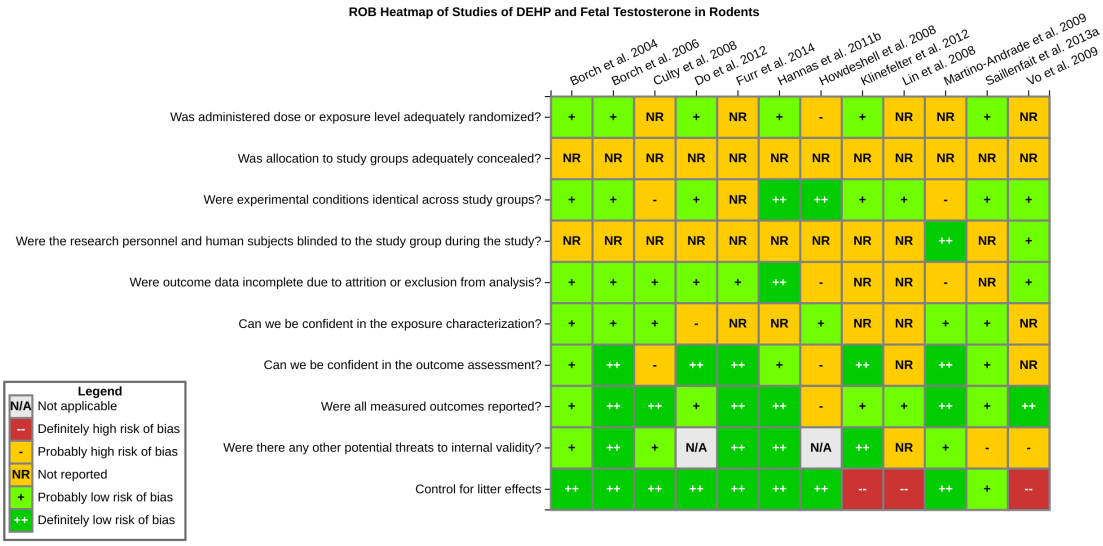


FIGURE 3-8 Risk of bias heatmap of studies of DEHP and fetal testosterone in rodents. In HAWC: https://hawcproject.org/summary/visual/362/.

- Factors potentially increasing confidence
  - Large magnitude of association or effect. Especially at higher doses, the effects on fetal testosterone could be considered large and robust, with overall summary estimates having z-scores of ≥7.0 and an overall summary estimate indicating >50% decreases. Moreover, these effect sizes were robust to multiple sensitivity analyses. Therefore, the meta-analysis supports an upgrade in confidence based on the large magnitude of effect.
  - Dose response. Upgraded because of strong evidence of dose response through meta-regression with statistically significant linear trends in either $\log_{10}$(dose) or dose. Moreover, these results were robust to multiple sensitivity analyses.

Table 3-5 presents the overall confidence ratings for the body of evidence on DEHP and fetal testosterone in rodents, and the details about how the ratings were determined are presented in Appendix C, Section C-4. There is high confidence in the body of evidence from experimental studies in animals.

Level of Evidence in the Health Effect. A meta-analysis performed on studies on DEHP and fetal testosterone (see "Meta-Analysis of DEHP and Alterations in Fetal Testosterone in Rats" presented later in the chapter) found consistent evidence of a decrease in fetal testes testosterone after in utero exposure to DEHP in rats. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DEHP is associated with a reduction in fetal testosterone in rats.

Hypospadias

Summary of the Evidence. Nine studies of DEHP and hypospadias were available. Of these studies, eight used the rat model and one used the mouse model (see Table 3-6). The exposure paradigm for all studies included the entirety of the male programming window. Animal hypospadias data were collected on a litter and/or an individual animal basis, and both methods were considered in the analysis. The hypospadias detection method for all rat studies was visual inspection of the phallus during postnatal life. Unlike the rat studies, the mouse study used a unique assessment methodology (urethral casting) and examined the phallus during fetal life (gestation day 19) (Liu et al. 2008).

TABLE 3-5 Profile of the Confidence in the Body of Evidence on DEHP and Fetal Testosterone Concentrations in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	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
DEHP	High (1 mouse,[a] 11 rat[b])	—	—	—	—	—	↑	↑	—	—	—	High

[a]Do et al. (2012).
[b]Borch et al. (2004, 2006); Culty et al. (2008); Howdeshell et al. (2008); Lin et al. (2008); Martino-Andrade et al. (2009); Vo et al. (2009); Hannas et al. (2011b); Klinefelter et al. (2012); Saillenfait et al. (2013a); Furr et al. (2014).

Risk of Bias Considerations. Figure 3-9 provides a summary of the risk of bias evaluation of the studies used by the committee to assess DEHP effects on hypospadias. The primary factors of concern for animal studies are reliability of outcome measure, blinding of researchers to treatment groups, and control for litter effects. The majority of studies did not adequately describe the method by which offspring were evaluated for hypospadias, and in 44% of the studies, blinding of the assessor was not reported. Most of the studies controlled for litter effects in the experimental design and/or statistical methods. The assessment 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. Because data reporting of methods and results was often incomplete, numerous studies received a "not reported" rating for one or more of these secondary risk of bias evaluations. There was no evidence of publication bias (see Appendix C, Section C-3).

Confidence in the Body of Evidence. The initial rating for the confidence in the animal studies was high because they involved controlled exposures, exposures occurred prior to outcome, outcomes were measured on individual animals, and a concurrent control comparison group was used (see Figure 3-2 for OHAT method for rating confidence). Confidence was downgraded because of the concern of significant risk of bias related to confidence in the outcome measure and blinding of investigators to the treatment groups. Confidence was also downgraded because of the concern of significant inconsistency in responses seen across studies. For example, for litters affected,[4] only one study was available in Sprague-Dawley rats (Saillenfait et al. 2009), which reported increased incidences of hypospadias (>30%) at the tested doses of 500 and 625 mg/kg-day, whereas among the three studies in Wistar rats, only one study reported effects, with a low incidence (9%) and only at a single intermediate dose (300 mg/kg-day). More studies reported effects as percent of animals affected, with all three studies in Sprague-Dawley rats reporting increased incidences (>10%, up to 100%), and two studies in Wistar rats reporting small increases in incidence (up to 5%) at an intermediate dose of 300 mg/kg-day (from a range up to 900 mg/kg-day).

TABLE 3-6 Summary of Animal Studies of DEHP and Hypospadias

Reference	Species/Strain	Doses, mg/kg-day	Route of Exposure	Window of In Utero Exposure	GD Definition in Study
Andrade et al. 2006	Wistar rat	0, 0.015, 0.045, 0.135, 0.405, 1.215, 5, 15, 45, 135, 405	Oral (gavage)	GD 6-PND 21	GD 0 = sperm positive
Christiansen et al. 2009	Wistar rat	0, 3, 15, 30	Oral (gavage)	GD 7-21	GD 1 = day after mating
Christiansen et al. 2010	Wistar rats: Study 1	0, 10, 30, 100, 300, 600, 900	Oral (gavage)	GD 7-PND 16	GD 1 = day after mating
	Wistar rats: Study 2	0, 3, 10, 30, 100	Oral (gavage)	GD 7-PND 16	GD 1 = day after mating
Gray et al. 2009	Sprague-Dawley rat	0, 11, 33, 100, 300	Oral (gavage)	GD 8-PND 17	GD 1 = sperm positive
Jarfelt et al. 2005	Wistar rat	0, 300, 750	Oral (gavage)	GD 7-PND 17	ND
Li et al. 2013	Sprague-Dawley rat	0, 500, 750, 1,000	Oral (gavage)	GD 12-19	GD 0 = vaginal plug
Liu et al. 2008	C57BL/6 mouse	0, 100, 200, 500	Oral (gavage)	GD 12-17	GD 0.5 = sperm positive
Saillenfait et al. 2009	Sprague-Dawley rat	0, 500, 625	Oral (gavage)	GD 12-21	GD 0 = sperm positive
Vo et al. 2009	Sprague-Dawley rat	0, 10, 100, 500	Oral (gavage)	GD 11-21	GD 0 = sperm positive

NOTE: BW, body weight; GD, gestation day; ND, not defined; PND, postnatal day.

[4]Because of possible litter correlation, effects reported as percent of litters affected are preferred over effects reported as percent of animals affected.

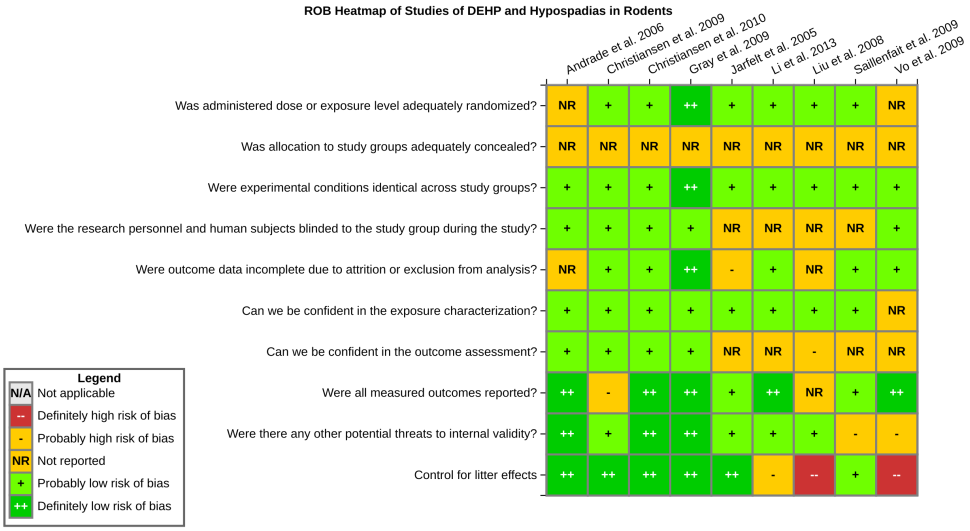


FIGURE 3-9 Risk of bias heatmap of studies of DEHP and hypospadias in rodents. In HAWC: https://hawcproject.org/summary/visual/360/.

Confidence in the body of evidence was upgraded because the background control incidence of hypospadias was reported as zero across all studies, so any positive finding was considered treatment related (i.e., rare outcome). Because hypospadias represents a dichotomous measure (present/absent) the presence of numerous studies reporting no incidence prevented the committee from completing a meta-analysis of the animal hypospadias data.

Table 3-7 presents the confidence ratings for the body of evidence on DEHP and hypospadias in rodents, and the details about how the rating were determined are presented in Appendix C, Section C-4. Overall there is moderate confidence in the body of evidence for hypospadias in animals.

Level of Evidence in the Health Effect. As described above, there is evidence of increased incidence of hypospadias in rats after fetal exposure to DEHP. Using the OHAT method (see Figure 3-3), a moderate confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a moderate level of evidence that fetal exposure to DEHP is associated with an increased incidence of hypospadias in male rats. The data suggest that Sprague-Dawley rats might be more sensitive to DEHP than Wistar rats.

TABLE 3-7 Profile of the Confidence in the Body of Evidence on DEHP and Hypospadias in Animals

		Factors Decreasing Confidence "—" If No concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	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
DEHP	High (1 mouse,[a] 8 rat[b])	↓	↓	—	—	—	—	—	—	—	↑	Moderate

[a]Liu et al. (2008).
[b]Jarfelt et al. (2005); Andrade et al. (2006); Christiansen et al. (2009, 2010); Gray et al. (2009); Saillenfait et al. (2009); Vo et al. (2009); Li et al. (2013).

Meta-Analyses of Animal Data

Meta-Analysis of DEHP and Reductions in AGD in Rats and Mice

The animal database for AGD and DEHP was judged to be amenable for meta-analysis. A summary of the analysis is provided below, and supporting details are presented in Appendix C, Section C-5. (Meta-analyses of studies on AGD and other phthalates are provided in Appendix C, Section C-6.) The following exclusions and groupings of studies were made to focus the analysis:

- Rat and mouse data were analyzed separately, due to known anticipated species differences in sensitivity (Johnson et al. 2012). Additionally, the rat data were subjected to a subgroup analysis by strain because of anticipated differential sensitivity across strains (Wilson et al. 2007).
- Studies in which exposures did not cover the entire male programming window (GD 16-18 in rats) were excluded.
- In many cases multiple AGD measures were reported for the same experiment. The measures selected for meta-analysis were in the following order of priority:
  - For studies that reported AGD at multiple time points in the same animals, the earliest postnatal time point was used.
  - For studies that reported AGD in multiple units, the order of preference was: AGD in mm/cube root of body weight, AGD in mm/body weight, and AGD in mm.

In all, 13 of the 16 rat studies and all three of the mouse studies were included in the analysis; three of the rat studies were excluded because they were missing group size values (Borch et al. 2004; Vo et al. 2009; Jones et al. 2015). Effect sizes were calculated as the log ratio of the mean difference between the treatment group and the concurrent control, multiplied by 100 ($y = 100 \times \ln$ [mean of treated group ÷ mean of control group]). For small changes, this is approximately equal to the percent change, but the resulting confidence interval is more symmetric and closer to normal (Hedges et al. 1999; Lajeunesse 2011). This normalization allows for treatment groups to be compared across studies and experiments. When normalized in this way, however, treatment groups within a study are correlated. Therefore, in one of the sensitivity analyses, effects were estimated using only the highest treatment group from each study. Additional sensitivity analyses were performed by sequentially excluding each study (all treatment groups for that study) (see Appendix C, Table C5-2).

Both an overall effect of any treatment and coefficients of meta-regressions were estimated. For meta-regressions, three models were used: a linear model in $y = a + b*\log_{10}(\text{dose})$ to test for a dose-response trend; a linear model $y = b*\text{dose}$; and a linear-quadratic model $y = b*\text{dose} + c*\text{dose}^2$ to model the dose-response shape. In the linear and linear-quadratic model, the intercept was omitted because the effect measures were already normalized relative to control levels. Additionally, for these models, the coefficients were rescaled in terms of the change per 100 mg/kg-day (e.g., $y = b*[\text{dose}/100] + c*[\text{dose}/100]^2$) for ease of interpretation. In all cases, random effect models were used, as described in the protocol. All analyses utilized random effects models, as implemented in the R "metafor" package. Sensitivity analyses included leaving one study out at a time and using only the highest dose group in each study (see Appendix C, Table C5-2). Benchmark dose estimates were calculated for an effect size of 5% (BMD_5; see Appendix C, Tables C5-3 and C5-5). The BMD_5 was calculated using the linear or linear-quadratic model, with the model selection based on the lowest AICc (Akaieke information criterion corrected for small sample size). The BMD_5 was calculated only for the "fixed effect"—that is, the estimated mean response across studies.

The results in rats for AGD are as follows:

- Statistically significant overall effect of a reduction in AGD (–3.96 [95% confidence interval (CI)]: –5.07, –285) and linear trends in $\log_{10}$(dose) (–1.97 [95% CI: –2.98, –0.96]) and dose (–1.55 [95% CI: –1.86, –1.24]). The overall effect was robust to leaving out individual studies.
- Under the linear-quadratic model, there was low heterogeneity (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 statistically 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 AICc, are shown in Figure 3-10.

The results in mice for DEHP and changes in AGD are as follows:

- No statistically significant overall effect, but statistically significant linear trends in $\log_{10}$(dose) (–1.77 [95% CI: –2.71, –0.83]) and dose (–2.03 [95% CI: –3.51, –0.55]).
- Under the linear-quadratic model (–5.71 [95% CI: –7.15, –4.27]), there was low heterogeneity (0%, p = 0.19), with the BMD_5 estimated to be 110 mg/kg-day (95% CI: 90, 150).
- When analysis was restricted to the highest dose group, there remained no statistically significant overall effect, and there was no longer a statistically significant linear trend.
- Overall effect was no longer statistically significant when leaving out some individual studies during the sensitivity analyses.

The results for the overall effect estimate, which had the lowest AICc, are shown in Figure 3-11.

Overall there is consistent evidence of a decrease in AGD in male rats after fetal exposure to DEHP, with a modest dose-response gradient. After fitting a meta-regression linear-quadratic model, heterogeneity was low or not detectable. In rats, the effects were robust to sensitivity analyses (see Appendix C, Table C5-2), involving removal of individual studies and use of only the highest exposure group. In mice, the effect estimates were similar after removing individual studies or restricting to the highest dose group, but in some cases, they lacked statistical significance owing to larger confidence intervals (see Appendix C, Table C5-4). Sprague-Dawley rats are less sensitive than Wistar rats, with a BMD_5 of around 300 mg/kg-day compared to 150 mg/kg-day. Mice have a BMD_5 of 250-350 mg/kg-day, which is in the range of the Sprague-Dawley rat.

Meta-Analysis of DEHP and Alterations in Fetal Testosterone in Rats

The animal database for DEHP and alterations in fetal testosterone was amenable to meta-analysis. A summary is provided below, and supporting details are presented in Appendix C, Section C-5. The same meta-analysis approach that was used to evaluate AGD was also applied to studies in rats of DEHP and fetal testes testosterone (there was only one study in mice). The same exclusions/groups were made, with the additional consideration that effects reported at least 6 h after dosing in acute studies were preferred over effects reported at earlier times because an effect should be greater at later time points. In all,

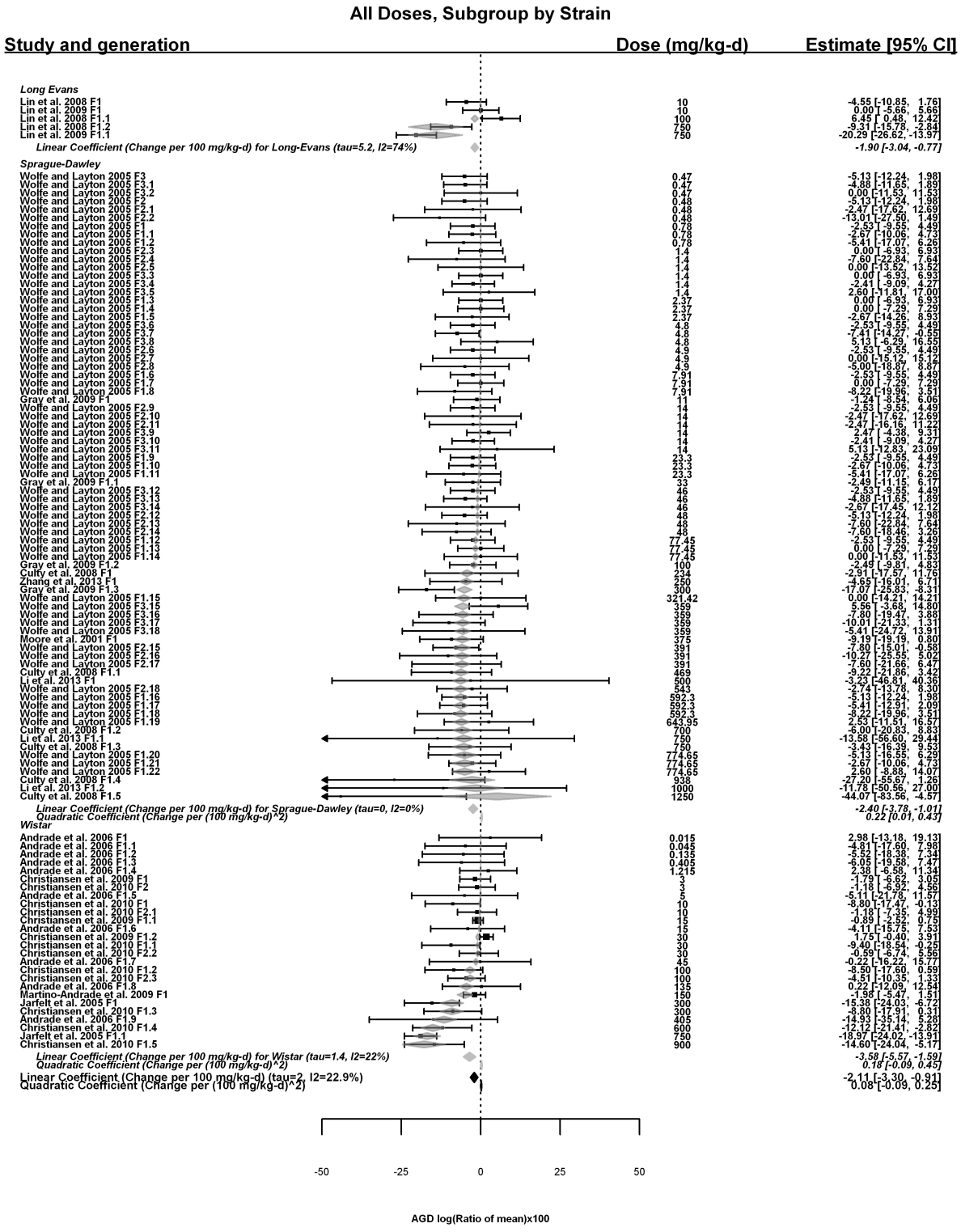


FIGURE 3-10 Results of the meta-regressions of studies on DEHP and AGD in rats. The overall effect of treatment in each strain is shown at the bottom of each subgroup analysis above as the change per 100 mg/kg-day.

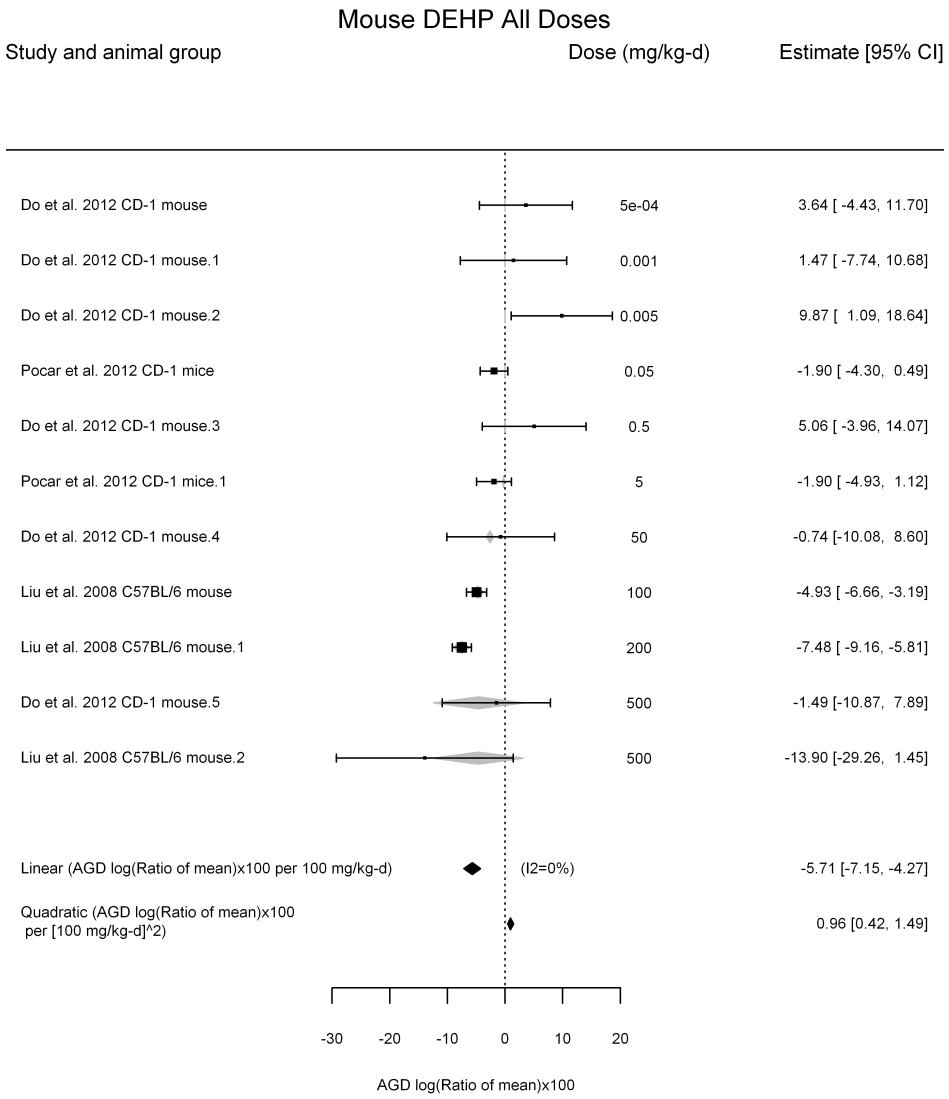


FIGURE 3-11 Results of the meta-analysis of studies on DEHP and AGD in mice. The overall effect of treatment is shown at the bottom of the figure as the change per 100 mg/kg-day.

7 of the 11 rat studies were ultimately included. The studies by Borch et al. (2004, 2006) and Vo et al. (2009) were excluded because they had missing group size values, and the study by Klinefelter et al. (2012) was excluded because testosterone measurements were taken after stimulation of the testes with luteinizing hormone. Benchmark dose estimates for effect size of 5% (BMD5) and 40% (BMD40) were calculated (see Appendix C, Table C5-8).[5] The results are as follows:

- Statistically significant overall effect (–110.14 [95% CI: –136.73, –83.54]) and linear trends in $\log_{10}$(dose) (–132.83 [95% CI: –171.03, –94.63]) and dose (–23.01 [95% CI: –26.24, –19.72]), with an overall effect that is large in magnitude (>50% change). The overall effect was robust to leaving out individual studies.

[5]BMD_{40}s were calculated for this end point because previous studies have shown that reproductive-tract malformations were seen in male rats when fetal testosterone production was reduced by about 40% (Howdeshell et al. 2015; Gray et al. 2016).

- Under the linear-quadratic model (–34.23 [95% CI: –47.02, –21.44]), there remains a substantial, statistically significant heterogeneity (I^2 >95%, p <0.001), with the BMD_5 estimated to be 15 mg/kg-day (95% CI: 11, 24). The BMD_{40} was found to be 160 mg/kg-day (95% CI: 120, 240).
- When analysis was restricted to the highest dose group, a larger overall effect, a larger linear trend in $\log_{10}$(dose), a consistent linear trend in dose, and consistent benchmark dose estimates were found.
- In subgroup analyses, there were statistically significant overall effects and linear trends in $\log_{10}$(dose) and dose for Sprague-Dawley and Wistar rats separately. Heterogeneity was reduced among Wistar rats (I^2 = 21%), but not among Sprague-Dawley rats (I^2 >95%).
- In subgroup analyses, there were statistically 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 to be slightly more sensitive than Wister rats, with slightly larger overall effect size and trend in $\log_{10}$(dose) and slightly lower benchmark dose estimates. The BMD_5 for Sprague-Dawley rats was estimated to be 13 mg/kg-day (95% CI: 9, 23), whereas the BMD_5 for Wistar rats was estimated to be 23 mg/kg-day (95% CI: 21, 24). The corresponding BMD_{40} estimates were 140 mg/kg-day (95% CI: 100, 230) for Sprague-Dawley rats and 230 mg/kg-day (95% CI: 210, 240) for Wistar rats.

The results of meta-regressions are shown in Figure 3-12 (linear-quadratic for Sprague-Dawley and linear for Wistar).

Overall there is consistent evidence of a decrease (>50% change) in fetal testes testosterone after DEHP treatment, with a strong dose-response gradient. Even after subgrouping by strain and meta-regression with dose, however, substantial heterogeneity remained in Sprague-Dawley rats. All three strains are outbred, so some of the residual heterogeneity may be due to genetic diversity. Nonetheless, the effects were robust to sensitivity analyses involving removal of individual studies and use of only the highest exposure group. Based on benchmark dose estimates, Sprague-Dawley rats are slightly more sensitive to these effects than Wistar rats (in contrast to the case with AGD).

Human-Health Effects Results on DEHP

Effects on AGD

Summary of the Evidence. The most robust data sets on DEHP were on AGD as measured by either AGD (ap [anopenile]) or AGD (as [anoscrotal]). The six epidemiologic studies that examined the relationship between biomarkers of DEHP exposure and AGD (ap or as) outcomes were all prospective cohort studies that enrolled pregnant mothers and their infants (see Table 3-8). A study by Suzuki et al. (2012) calculated and reported AGD index rather than AGD.

The cohort studies were performed in the United States (Swan 2008; Swan et al. 2015; Martino-Andrade et al. 2016); Scandinavia (Bornehag et al. 2015; Jensen et al. 2016); and Mexico (Bustamante-Montes et al. 2013). The studies varied in timing of when urinary phthalate metabolites were measured during pregnancy, age when infant AGD was measured, and the reporting on reliability of AGD measurements. Jensen et al. (2016) and Bustamante-Montes et al. (2013) measured urinary phthalate metabolites in the third trimester only, whereas Swan (2008) measured them throughout pregnancy but on average late in pregnancy. Bornehag et al. (2015) measured urinary phthalate metabolites during the first trimester of pregnancy.

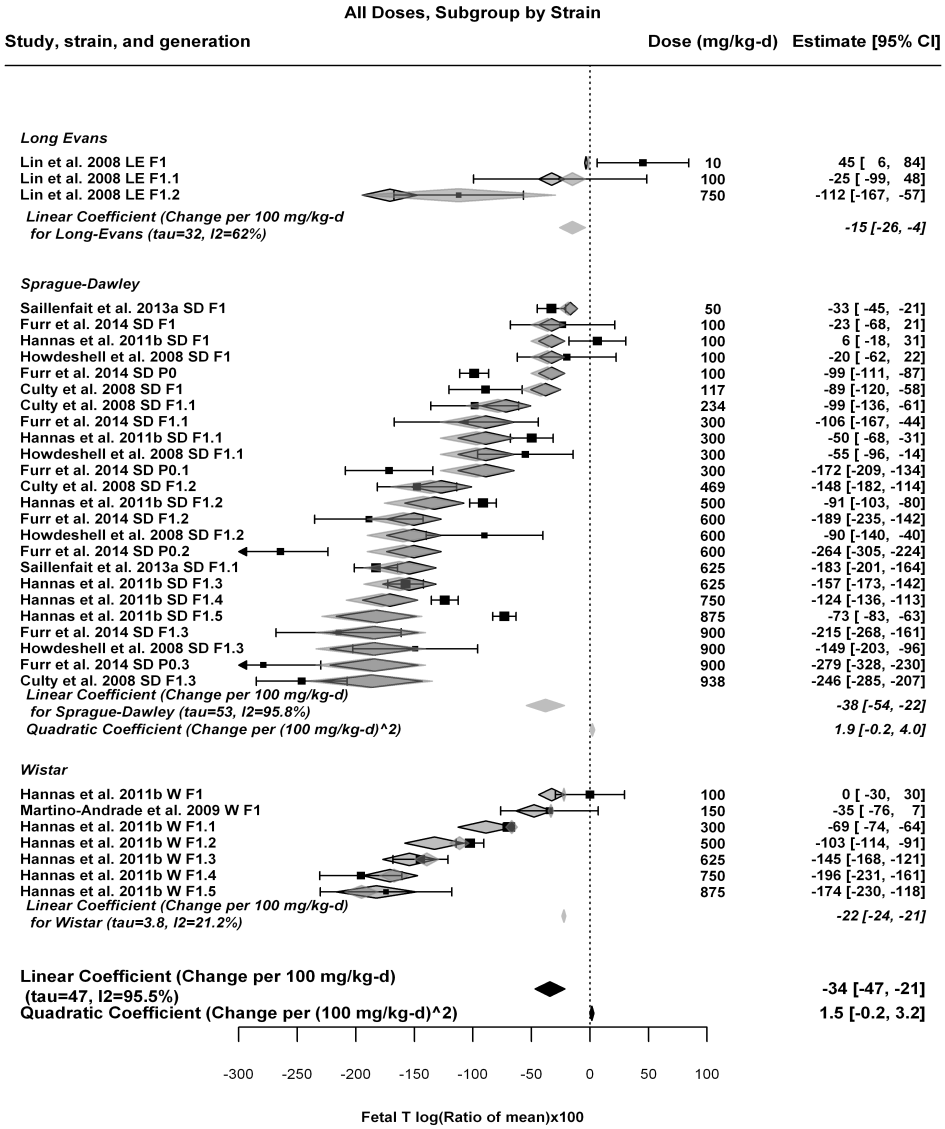


FIGURE 3-12 Results of the meta-regressions of studies on DEHP and fetal testosterone in different strains of rat. The overall effect of treatment for each strain is shown at the bottom of each subgroup analysis as the change per 100 mg/kg-day.

The results of a fifth study from the same cohort were reported in two publications: Swan et al. (2015) used measurements in women who were less than 13 weeks pregnant, and Martino-Andrade et al. (2016) used measurements of second and third trimester urinary phthalate metabolites. AGD measurements were performed on infants up to 36 months of age in the studies by Swan (2008) and Bornehag et al. (2015), whereas measurements were taken in infants who were 3 months old or younger in the studies by Bustamante-Montes et al. (2013), Swan et al. (2015), Jensen et al. (2016), and Martino-Andrade et al. (2016). Reports of AGD measurement reliability varied across studies, as some included data on intra- and inter-rater reliability and others did not. (Appropriate methods involve standardized training for all examiners using calipers as the primary measurement instrument and continued repeat measurements on the same subject as well as by different trained examiners on the same subject throughout the study to

ensure low intra-rater and inter-rater variability.) No study was excluded on the basis of failing to report rater reliability, however, as long as measurement methods were appropriate and well described. All studies used state-of-the-art analytical chemistry methods to measure urinary phthalate metabolites, and they included collection of and adjustment for important potential confounding variables such as measures of infant body size and maternal demographic factors.

Risk of Bias Considerations. The risk of bias ratings for the individual studies are presented in Figure 3-13. The primary factors of concern for human studies are confounding, exposure characterization, and outcome assessment (including blinding of outcome assessors). The questions used to evaluate risk of bias in the individual studies are provided in Appendix D, Section D-1e. For this data set, the studies had either a low or a very low risk of bias in these domains. Specifically, the risk of bias assessment for urinary phthalate metabolite measurements (biomarkers of exposure) considered the reliability of the test methods (e.g., use of high performance liquid chromatography with tandem mass spectrometry) and whether the exposure biomarker was assessed in a relevant time-window for development of the outcome. It also considered whether there was a measure of urinary dilution that was accounted for in the analysis, such as urinary specific gravity or creatinine. The short half-life of DEHP (<24 h) (Koch et al. 2004) may contribute to exposure misclassification, an important issue in environmental epidemiologic studies. The gold standard would be multiple 24-h urine samples during the relevant sensitive window of exposure, which is difficult to obtain in human studies. Although one spot urine sample is not the best measure of long-term exposure over the relevant prenatal programming period, exposure misclassification would likely introduce random noise and bias toward the null.

Risk of bias evaluation of the outcome assessment considered the methods for determining the outcome, whether the outcome had been assessed consistently across all groups, and whether the outcome assessors had been blinded to the study groups or exposure levels prior to assessing the outcomes. Given the study designs for the epidemiologic studies that examined AGD, it was unlikely that examiners measuring AGD would know urinary phthalate levels at the time the AGD measurement was made.

Risk of bias assessment of the studies of DEHP and AGD also included assessment for important confounding variables such as age, race/ethnicity, weight/body size, and age at exam. Most studies measured multiple urinary phthalate metabolites in addition to DEHP metabolites. There can be a correlation among phthalate metabolites from different diesters. This may contribute to confounding by other metabolites. There was no evidence of publication bias (see Appendix D, Table D3-3).

Confidence in the Body of Evidence. The initial rating for the confidence in the human studies was moderate based on the following three criteria: exposures occurred prior to outcome, outcomes were measured on individuals, and a (control) comparison group was used (see Figure 3-2 for OHAT method for rating confidence). A meta-analysis of these data is presented later in this chapter, and it provides additional information concerning the confidence ratings. Specifically, meta-analysis supports the following:

- Factors potentially decreasing confidence
  - Unexplained inconsistency. No downgrade is warranted because the meta-analysis I^2 statistic was 0%. In some cases, larger values (up to 54%) were estimated in sensitivity analyses, but these are given less weight because they involved the use of less preferred outcome or exposure estimates, which are expected to introduce more heterogeneity.
  - Imprecision. The meta-analysis also supports that imprecision in the results is not a concern; the summary estimate has a 95% confidence interval of –6.49, –1.66, and the confidence intervals for the sensitivity analyses were similar. Therefore, the same causal conclusion would be reached based on either end of the confidence interval. As discussed in the GRADE framework (Guyatt et al. 2011), confidence intervals that would result in different conclusions depending on whether the upper or lower limit is used can result in a downgrade due to imprecision. Because the summary estimate and its statistical significance is robust to multiple sensitivity analyses, the meta-analysis would support the conclusion that imprecision is not a serious concern.

TABLE 3-8 Summary of Human Studies of DEHP and AGD

Reference	Study Design, Location, Years	Sample Size	Metabolites Measured in Maternal Urine, Time of Measurement	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	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	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	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 weeks gestational weeks	AGD (as) AGD (ap) At delivery AGD index used in analyses	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; MEHP; 5OH-MEHP; 5oxo-MEHP; 5carboxy-MEPP; all three trimesters	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

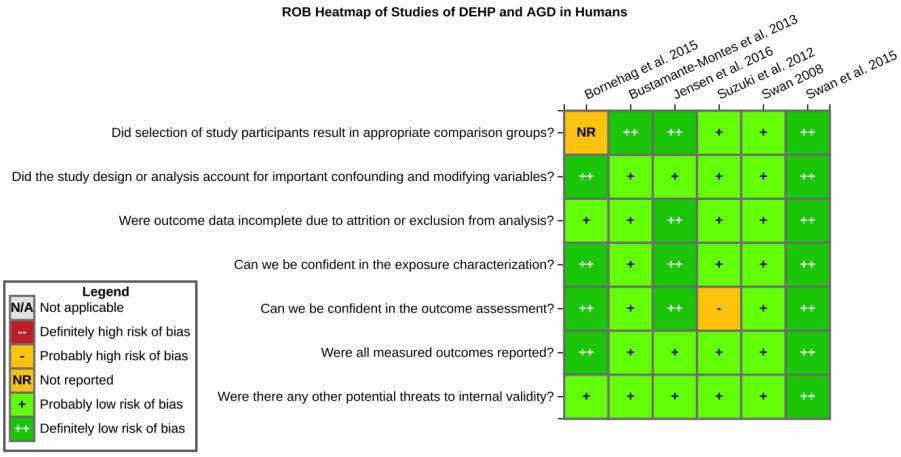


FIGURE 3-13 Risk of bias heatmap of studies of DEHP and AGD in humans. The study by Martino-Andrade et al. (2016) does not appear in the heatmap because it is linked to the Swan et al. (2015) study; it has the same risk of bias evaluation as that study. In HAWC: https://hawcproject.org/summary/visual/341/.

- Factors potentially increasing confidence
  - Large magnitude of association or effect. An upgrade is not warranted. Although the effect size observed for AGD of a 4% decrease per 10-fold increase in DEHP metabolite concentration can be considered relatively large—as this degree of change in AGD in experimental animal studies is associated with around a 40% decrease in fetal testosterone production—the smaller end of the confidence interval is an effect size of –1.66%, and in some of the sensitivity analyses, the smaller end of the effect size is <1%. Therefore, a small magnitude of effect cannot be ruled out with reasonable confidence.
  - Dose response: The effect estimates of AGD are estimates of slopes; thus, they are based on the assumption of a monotonic dose-response relationship between exposure and effect. One study reported dose-response information independent of slope estimates, and it was not informative due to wide confidence intervals. Therefore, the meta-analysis would not support an upgrade in the confidence conclusion based on evidence of a dose-response gradient.

There were no changes in the confidence rating for the human evidence after considering factors that could increase or decrease confidence. Table 3-9 presents an evidence profile of the findings on DEHP and AGD in humans, and additional details about how the moderate rating was determined is presented in Appendix D, Section D-3.

Level of Evidence in the Health Effect. The results show a consistent pattern of findings that higher maternal urinary concentrations of DEHP metabolites during pregnancy (during the prenatal male genital programming window) are associated with a smaller AGD in male infants compared to infants whose mothers had lower DEHP exposures during pregnancy. Consistent reductions in AGD were found across multiple studies; the small amount of heterogeneity observed may be due to sample size differences, AGD measurement variability, urinary metabolite concentration variability, and the potential for residual confounding. A meta-analysis (presented later in this chapter) found consistent evidence of a decrease in AGD being associated with increasing urinary concentrations of the sum of DEHP metabolites. Using the OHAT method (see Figure 3-3), a moderate confidence rating in the body of evidence (described above) and evidence of an effect result in a conclusion that there is a moderate level of evidence that fetal exposure to DEHP is associated with a reduction in AGD.

TABLE 3-9 Profile of the Confidence in the Body of Evidence on DEHP and AGD in Humans

			Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence			
Phthalate	Metabolite(s)	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	FINAL CONFIDENCE RATING
DEHP	MEHP; 5-oxo-MEHP; 5OH-MEHP; sumDEHP metabolites	Moderate (6 prospective)[a]	—	—	—	—	—	—	—	—	Moderate

[a]Swan et al. (2008); Bustamante-Montes et al. (2013); Bornehag et al. (2015); Swan et al. (2015); Jensen et al. (2016); Martino-Andrade et al. (2016).

Testosterone Concentrations Measured During Gestation or at Delivery

For fetal testosterone assessment, only three human studies met the criteria for inclusion of having measurements of urinary concentrations of DEHP metabolites, either sumDEHP metabolites or individual DEHP metabolites. Another study was excluded from consideration because the authors measured only phthalate monoester metabolites (MEHP) in maternal blood. Because of concern with external contamination in matrices such as blood (Calafat et al. 2015), studies that measured only monoester phthalate metabolites in blood were excluded. Of the three studies, one study measured testosterone concentrations in amniotic fluid (Jensen et al. 2015), one measured testosterone in cord blood (Lin et al. 2011), and the other measured maternal serum testosterone concentrations during pregnancy (Sathyanarayana et al. 2014). The Jensen et al. (2015) study utilized a large biobank of amniotic fluid samples collected in Denmark between 1980 and 1996 to study the cross-sectional association of amniotic fluid concentrations of testosterone with oxidative metabolites of DEHP (and DINP). Within this biobank, they conducted a nested case-control study with 270 cases of cryptorchidism, 75 cases of hypospadias, and 300 male controls. Among the controls there was no association of amniotic fluid concentrations of mono-(2-ethyl-5-carboxypentyl) phthalate (MECPP) with testosterone concentrations. The Lin et al. (2011) study measured maternal urinary concentrations of DEHP metabolites and found no association with cord serum concentrations of testosterone among male newborns. The Sathyanarayana et al. (2014) study was not further considered because it is not known if maternal serum concentrations directly reflect fetal testosterone production; therefore, it would not be possible to make inferences about the association of urinary DEHP metabolites with fetal testis testosterone production.

Based on the disparate matrices used to estimate fetal testis testosterone production (amniotic fluid or cord blood), the differences in timing of measurement of testosterone (during pregnancy or at delivery), and the dearth of studies, the committee determined that the data were inadequate to draw any conclusions. The committee also recognized that human studies on fetal testosterone production in relation to phthalate exposure are logistically very difficult to conduct, and it is not possible to directly determine fetal testis testosterone production. Although the Jensen et al. (2015) study used a design that most closely approximates assessing fetal testis testosterone production during pregnancy, its interpretation was hindered by uncertainty regarding relevance of amniotic fluid levels of testosterone to fetal testis testosterone production: namely, the pharmacodynamics of testosterone levels within amniotic fluid. Furthermore, uncertainty regarding the pharmacokinetics of fetal phthalate metabolism limited the interpretation of MECPP concentrations in amniotic fluid.

Hypospadias

For hypospadias, two human studies (Chevrier et al. 2012; Jensen et al. 2015) were available on measures of DEHP metabolites. Jensen et al. (2015) conducted a case-control study on hypospadias that was nested within a large biobank of amniotic fluid samples collected in Denmark between 1980 and 1996. They measured amniotic fluid MECPP concentrations among 75 cases of hypospadias and 300 controls. There was no association of MECPP with odds of hypospadias. The Chevrier et al. (2012) study, a nested case-control study with 21 cases of hypospadias, did not find an association between hypospadias and urinary concentrations of DEHP metabolites. These two studies used disparate matrices to measure DEHP metabolites (amniotic fluid and urine) and were very small in size, limiting their power. Given that these were the only two studies, the committee determined that the data were inadequate to draw any conclusions or conduct a meta-analysis.

Meta-Analysis of Human Data on AGD and DEHP

The epidemiologic database on ADG and DEHP was judged to be amenable for meta-analysis. Five cohorts contributed data to the analysis (Swan 2008; Bustamante-Montes et al. 2013; Bornehag et al. 2015; Swan et al. 2015; Jensen et al. 2016). The preferred measures for each study were:

- Outcome: AGD (as) is preferred over AGD (ap) because it is a more reliable measurement.
- Time of exposure measurement: The first trimester is preferred over the second trimester, which is preferred over the third trimester, because the male programming window is in the first trimester.
- Exposure metric: Sum of DEHP metabolites is preferred over MEHP, which is preferred over any of the other DEHP metabolites, because the sum better reflects the parent compound exposure.

The primary study data, using the preferred DEHP exposure biomarkers and outcome measure for each study, are presented in Table 3-10, and the result of the analysis is presented in Figure 3-14. For the studies by Bustamonte-Montes et al. (2013) and Swan (2008), the CIs were estimated using the reported p-value, assuming a normal distribution. For other studies, confidence intervals were included in the published manuscript.

TABLE 3-10 Summary of Human Data Used in Meta-Analysis of DEHP and AGD

Study	Outcome	Exposure	Exposure metric	Beta estimate, %	Lower CI, %	Upper CI, %
Bornehag et al. 2015	AGD (as)	sumDEHP metabolites	Maternal urine	–2.80	–9.69	4.06
Bustamante-Montes et al. 2013	AGD (as)	MEHP	Maternal urine	–0.36	–10.31	9.59
Jensen et al. 2016	AGD (as)	sumDEHP metabolites	Maternal urine	–2.93	–8.43	2.55
Swan 2008	AGD (ap)	MEHP	Maternal urine	–4.98	–9.06	–0.89
Swan et al. 2015	AGD (as)	sumDEHP metabolites	Maternal urine (trimester 1)	–5.10	–9.70	–0.53

NOTE: AGD (ap), distance between the anus and base of the phallus; AGD (as), distance between the anus and scrotum.

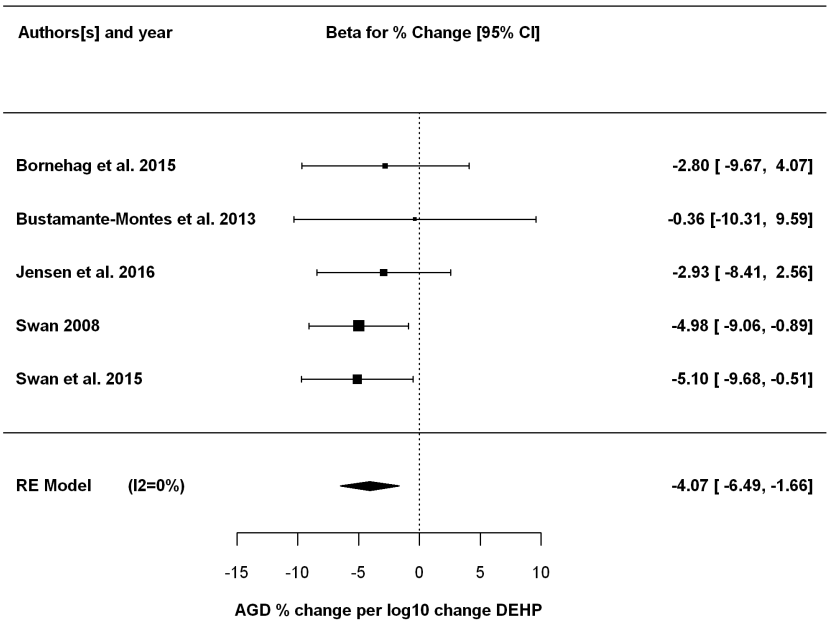


FIGURE 3-14 Results of the meta-analysis of studies on DEHP and AGD in humans are shown as the percent change per $\log_{10}$ change in DEHP concentration.

Slopes (beta coefficients) are reported in units of change in mm/$\log_{10}$ change in urinary concentrations of DEHP metabolites. Two factors a priori may affect comparability across studies. First, there are baseline differences in AGD (as) across different studies due to such demographic factors as age at measurement of AGD, which is affected by weight and body size. For instance, the mean AGD (as) reported by Bustamante-Montes et al. (2013) was 12.4 mm, whereas the mean AGD (as) reported by Bornehag et al. (2015) was 41.4 mm. Second, AGD (as) is shorter than is AGD (ap). For instance, in the study by Jensen et al. (2016), mean AGD (as) was 36.9 mm, whereas mean AGD (ap) was 70.2 mm. Therefore, the same change in distance may reflect different percentage change in AGD across studies in end points. 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 is standardized to a percent change in AGD per $\log_{10}$ change in urinary DEHP metabolite concentrations.

All analyses utilized random effects models, as implemented in the R metafor package. Sensitivity analyses included leaving one study out at a time; using alternative exposure and outcome measures for each study, one at a time; and restricting analyses to use of the same exposure measure (the sum of DEHP metabolites or MEHP) and/or the same outcome measure AGD (as) or AGD (ap). Figure 3-15 shows the sensitivity analyses that were performed by leaving one study out at a time.

In the primary analysis, five studies, with beta coefficients standardized to a percent change per $\log_{10}$ change in DEHP exposure, were analyzed using a random effects model. A statistically significant summary estimate of –4.07 (95% CI: –6.49, –1.66; [p = 0.001]) was found for the change in AGD per $\log_{10}$ increase in DEHP exposure. 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.

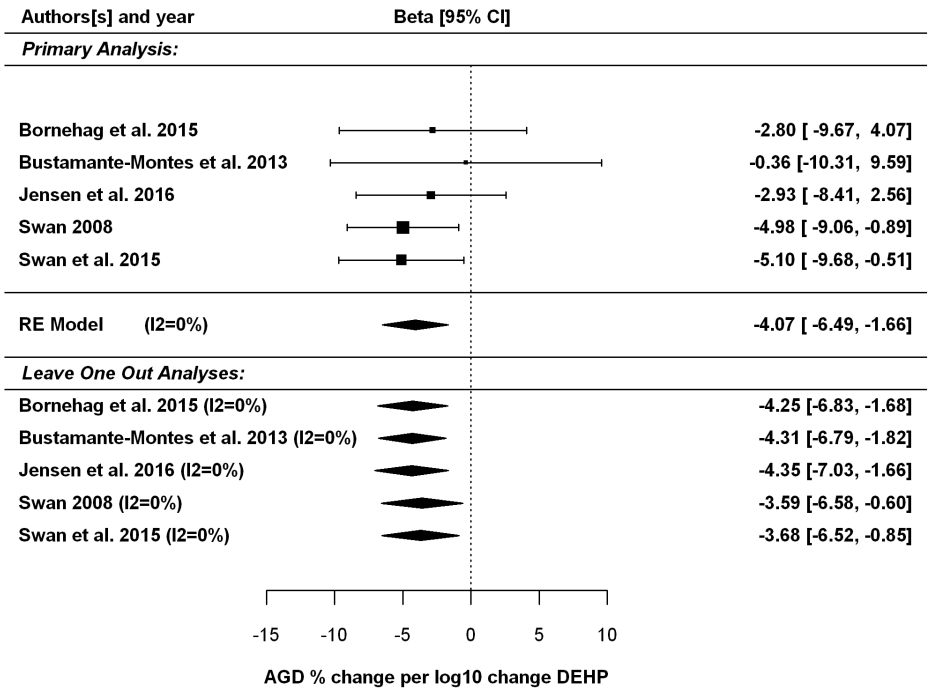


FIGURE 3-15 Results of the sensitivity analysis of the meta-regression of studies on DEHP and AGD in humans as shown as the percent change in AGD per $\log_{10}$ change in DEHP concentration. Analyses were performed leaving one study out at a time.

Leaving one study out at a time, the summary estimates ranged from –4.35 to –3.59. The summary estimate remained statistically significant in all cases, with p-values ranging from 0.001 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 from Jensen et al. (2016).

Sensitivity analyses were further performed using alternative effect estimates for each study (see Appendix D, Table D4-1). The summary estimates ranged from –4.78 to –1.51. In 11 of the 42 alternative analyses, the summary estimates were no longer statistically significant (summary estimates range from –1.51 to –2.69), with p-values ranging from 0.050 to 0.41. All of the nonstatistically significant alternative analyses involved replacing the Swan et al. (2015) results 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 statistically significant).

Finally, eight 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) (see Appendix D, Table D4-1). The resulting summary estimates ranged from –4.2 to –2.0, all of which were statistically significantly different from 0. Additionally, there was no observed heterogeneity in any of these cases ($I^2 = 0$).

Overall, there is consistent evidence of a decrease in AGD being associated with increasing urinary concentrations of the sum of DEHP metabolites and of magnitude around 4% for each $\log_{10}$ increase in DEHP concentrations. There was no evidence of heterogeneity in the primary analysis, and this result was robust to removing individual studies. The result was also robust to 50 additional sensitivity analyses that used alternative effect size estimates. In about 80% of these sensitivity analyses, the summary estimate remained statistically significant. Moreover, the eight sensitivity analyses involving stricter criteria for homogeneous exposure and outcome measures had summary measures that were statistically significant with no observed heterogeneity. The majority of the weight in the preferred summary estimate, however, is from two studies from different cohorts (i.e., independent study populations) with the same first author

(Swan 2008; Swan et al. 2015). Dropping both of these studies would result in a summary estimate that is consistent with all analyses (negative indicating a reduction in AGD with a $\log_{10}$ increase in DEHP) but was no longer statistically significant (not shown, –2.48 [95% CI: –6.42, 1.45], $I^2 = 0$). Overall, however, greater weight is given to the primary analysis because it includes all the available studies that met the prespecified inclusion criteria and because it reflects the preferred measures of outcome and exposure.

EVIDENCE INTEGRATION FOR AGD

Evidence synthesis for AGD was conducted in a three-part process. First, the confidence ratings for the human and animal studies were translated into conclusions about level of evidence of health effects using the procedure outlined by OHAT (performed earlier in this chapter). Second, an initial hazard identification conclusion was reached by integrating the conclusion about level of evidence for the human and the animal evidence streams. Third, the degree of support from mechanistic data was considered and discussed in reaching final hazard identification conclusions for AGD.

Initial Hazard Conclusion for AGD

As described in earlier sections, the level of evidence for fetal exposure to DEHP being associated with reductions in AGD was high for the animal evidence and moderate for the human evidence. Using the OHAT hazard identification scheme (see Figure 3-4), an initial hazard conclusion is reached that DEHP is presumed to be a reproductive hazard to humans. The human and animal bodies of evidence present a consistent pattern of findings that prenatal exposure to DEHP is associated with reduced AGD.

Consideration of Mechanistic Data on AGD

Mechanistic data available from the rat support that the following steps are involved in phthalate reproductive toxicity: (1) metabolism of the phthalate diester to the monoester; (2) decreased expression of genes that regulate cholesterol metabolism and steroidogenesis in fetal Leydig cells; (3) decreased production of fetal testis testosterone; and (4) reduced expansion of the perineum resulting in a change in AGD and altered urethral closure resulting in hypospadias (Howdeshell et al. 2015; see Figure 3-16). Several of these elements will be discussed in greater detail below.

Phthalate Metabolism and Pharmacokinetics

A critical first step in phthalate reproductive toxicity involves the hydrolysis of the diester phthalate to the more toxic monoester metabolite. Subsequent metabolic steps include formation of the MEHP-glucuronide by UDP-glucuronyl transferase and the creation of oxidized metabolites formed by cytochrome P450 4A that are further oxidized by alcohol or aldehyde dehydrogenases (Albro and Lavenhar 1989).

One of the more significant species differences in metabolism relates to the rate at which MEHP is formed by lipase. For example, lipase activity seen in mouse liver homogenates was markedly higher than that observed in marmosets (1,339 ± 261 pmol/g versus 62 ± 11 pmol/g) (Ito et al. 2005). Notable species differences in Vmax, Km, and Vmax/Km ratio were also seen between mice, rats, and marmosets, suggesting that species differences in lipase activity may result from different enzyme affinities and different expression levels of the enzyme (Ito et al. 2005). Follow-up studies have confirmed that humans, like marmosets, have lower hepatic lipase activity when compared with mice although the inter-individual variability is quite large in people (Ito et al. 2014). Other species differences in enzyme activity are seen with several of the other enzymes involved in DEHP metabolism; however, the extent of the difference is not as large as that seen with lipase (Ito et al. 2005).

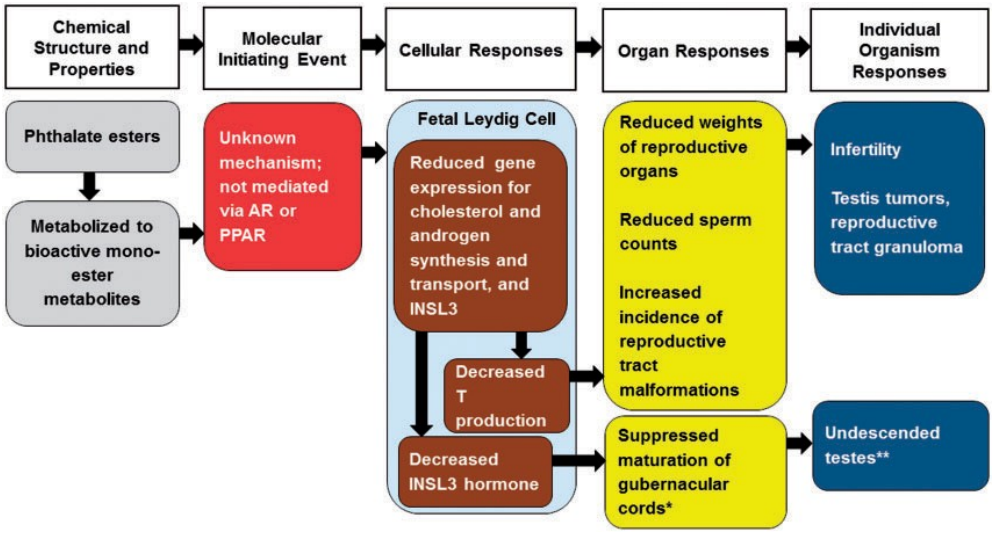


FIGURE 3-16 Theoretical steps involved in male reproductive toxicity following phthalate exposure during the in utero male programming window. *Suppressed maturation of the gubernacular cords contributes to incidence of cryptorchidism (undescended testes) due to role of gubernacular cords in transabdominal descent of the testes. **Undescended testes may contribute to incidence of testicular cancer. NOTE: AR, androgen receptor; INSL3, testis insulin-like 3 (INSL-3); PPAR, peroxisome proliferator active receptor; T, testosterone. SOURCE: Howdeshell et al. (2015). Reprinted with permission; copyright 2015, *Toxicological Sciences*.

The reproductive toxicity of a structurally related phthalate metabolite, monobutylphthalate (MBP), has been evaluated in pregnant marmosets. Fetal exposure of marmosets to MBP at 500 mg/kg-day during gestation weeks 7-15 did not affect plasma testosterone levels at birth or hypospadias or other changes in reproductive-tract development (McKinnell et al. 2009). Several marmosets exposed to DEHP in utero did develop clusters of undifferentiated testicular germ cells of unknown biological significance, however (McKinnell et al. 2009).

Additional pharmacokinetic studies have been performed in rodents, marmosets, and people. For example, Kessler et al. (2004) found that peak blood concentrations of MEHP in rats was approximately three times higher (range: 1.3-7.5 µg/mL) than in marmosets when both species were exposed similarly to DEHP. In addition to differences in metabolism there may also be differences in the way conjugated DEHP metabolites are eliminated. For example, several free oxidized DEHP metabolites are observed in the plasma of rats but not of marmosets (which more quickly glucuronidate these metabolites) following DEHP exposure (Kurata et al. 2012). Kessler et al. (2012) and Koch et al. (2004) evaluated the pharmacokinetics of DEHP in adult human volunteers after they ingested deuterium-labelled DEHP. A striking observation in the human pharmacokinetic study was that peak concentrations (C_{max}) and area under the curve (AUC) for MEHP and DEHP in human serum are much greater than those reported for either rats or marmosets given comparable administered doses. Kessler et al. (2012, p. 289) concluded that "the MEHP blood burden at a given DEHP dose per kg body weight will be higher in humans than in the animals." Similar studies have not been performed in pregnant women.

The pharmacokinetics of DEHP has also been investigated in chimeric mice transplanted with human hepatocytes, experiments that also supported development of a simplified physiologically based pharmacokinetic (PBPK) model for DEHP (Adachi et al. 2015). The PBPK model consists of gastrointestinal, liver, and central compartments. Resulting PBPK model predictions suggest that MEHP will be cleared from plasma in humans similarly to what is seen in mice, but fecal elimination of MEHP (and other oxidized metabolites) will have a higher rate in people (Adachi et al. 2015). An important caveat remains, however, that placental transfer of phthalates to the fetus is incompletely understood and described in this model.

Decreased Expression of Genes That Regulate Steroidogenesis

Studies have shown that in utero exposure of animals to DEHP can produce a reduced expression of proteins involved in steroidogenesis in the fetal testis. Affected proteins can include CYP11A1, CYP17A1, translocator protein (18-kDa), and STAR (Gray et al. 2000; Borch et al. 2006; Culty et al. 2008). Although changes in STAR and other enzymes are seen following in utero exposure, knowledge concerning the molecular initiating event involved in these reductions in enzyme activity remains unclear. This data gap is not addressed in current high-throughput assay systems (e.g., ToxCast) because the steroidogenic assay used in these programs often relies on a human adrenal cell line (Karmaus et al. 2016) and adrenal steroidogenesis in vivo is not affected by phthalate exposure via the same mechanism (i.e., decrease in steroidogenic pathway gene expression) (Thompson et al. 2005; Martinez-Arguelles et al. 2011).

Decreased Production of Fetal Testis Testosterone

The committee's systematic review found a high level of evidence that in utero exposure to DEHP in rats is associated with a reduction in fetal testosterone levels. Sprague-Dawley rats appeared to be slightly more sensitive to this effect than Wister rats. The BMD_5s for Sprague-Dawley and Wistar rats were estimated to be, respectively, 13 (95% CI: 9, 23) and 23 (95% CI: 21, 25) mg/kg-day. In contrast, based on the committee's BMD_5 estimates for AGD, Sprague-Dawley rats were approximately twofold less sensitive to DEHP-induced effects on AGD when compared with Wistar rats, suggesting possible strain differences in the quantitative relationship between decreases in fetal testosterone and changes in AGD.

Studies have shown 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), have shown that reduced fetal testicular testosterone occurs in rats, but not in mice, following in utero exposure (Gaido et al. 2007; Johnson et al. 2012).

Biological Plausibility

The mechanistic data developed in vitro and in animal models provide evidence that the DEHP effects on AGD in humans identified by the committee's systematic review are biologically plausible. Moreover, androgen-dependent development of the male reproductive tract and androgen-dependent AGD appear to be well conserved across mammalian species (including humans). Nevertheless, the mechanistic data were not sufficient to result in an upgrade in the committee's final hazard identification for AGD (see Figure 3-4).

Final Hazard Conclusion on AGD

On the basis of the committee's evidence integration of the animal and the human evidence on DEHP and effects on AGD and consideration of relevant mechanistic data, the committee concluded that DEHP is presumed to be a reproductive hazard to humans.

EVIDENCE INTEGRATION FOR FETAL TESTOSTERONE

The approach to integrating the animal and the human evidence on the effects of DEHP on fetal testosterone was the same as that used for AGD.

Initial Hazard Conclusion for Fetal Testosterone

As described in earlier sections, the level of evidence for fetal exposure to DEHP being associated with decreased fetal testosterone synthesis was high for evidence in rats and was inadequate for evidence

in humans. Using the OHAT hazard identification scheme (see Figure 3-4), an initial hazard conclusion was reached that DEHP is presumed to be a reproductive hazard to humans.

Consideration of Mechanistic Data

Decreased Testosterone Following DEHP Exposure

As mentioned earlier, mechanistic data available on the rat support the hypothesis that decreased production of fetal testis testosterone occurs in animals following fetal exposure to DEHP. The decrease in production in rats may be secondary to reduced expression of STAR and other proteins involved in fetal testis steroidogenesis (Gray et al. 2000; Borch et al. 2006; Culty et al. 2008). These in vivo studies are supported by some in vitro studies that demonstrated decreased testosterone production in cultured rat fetal testes exposed to MEHP but not the parent phthalate DEHP (Chauvigné et al. 2009). Other studies conducted with cultured rat testes failed to show an effect of MEHP at in vitro concentrations up to 10 μM (Stroheker et al. 2006). Likewise, these effects have not been observed in cultured human fetal testes treated with MEHP (Lambrot et al. 2009). Pharmacokinetic studies of DEHP and other phthalates suggest that differences in decreased fetal testes testosterone production reflect differential potency for testosterone inhibition rather than differences in tissue dosimetry (Clewell et al. 2010).

Another line of mechanistic data the committee considered was the result of xenograft studies performed in rodents for exploring differences in species sensitivity. In one of these experiments, fetal rat, mouse, and human testes were implanted in nude rats or mice exposed to DBP at 250 or 500 mg/kg-day for 1-3 days (Heger et al. 2012). Only rat xenografts exhibited statistically significant decreases of steroidogenic gene expression (including *Star*) and testosterone secretion (Heger et al. 2012). As with implanted mouse testes, human testes did not develop statistically significant phthalate-induced suppression of steroidogenic gene expression (including *Star*) following host exposure up to 500 mg/kg-day for two days (Heger et al. 2012). Similar results have been obtained in experiments that examined the responses of human fetal tissues (collected at gestational weeks 10-23) implanted in castrated immunodeficient mouse hosts for 6 weeks (Mitchell et al. 2012). Qualitatively, the data indicate that human testes appear to be less sensitive to the effects of phthalates than the rat testes. A limitation of those studies is that they rely on small numbers of human tissue samples. Consequently, the committee performed a literature search on February 9, 2017, and identified two studies that provide xenograft data regarding testosterone concentrations (Mitchell et al. 2012; Spade et al. 2014). Mean and standard errors were digitized and standard errors were converted to standard deviations. The effect measure is the log ratio of the mean between treated and control, times 100 (which for small values is close to the percent change). Random effects models were fit for overall effect. There were too few studies to do sensitivity analyses. The overall effect size was estimated to be –15.7 (95% CI: –51.8, 20.4), corresponding to a percent change of –14.5% (95% CI: –40.4, 22.6). There was no heterogeneity observed ($I^2 = 0\%$). While a trend toward decreased serum testosterone was observed, it was not statistically significant. Due to the low precision of the estimate, however, the data are inadequate to conclude whether an effect may have occurred, since they are consistent with effect sizes ranging from a 40% decrease to a 23% increase in serum testosterone. Figure 3-17 presents the results of the meta-analysis.

Biological Plausibility

The mechanistic data developed in vitro and in animal models provides evidence that DEHP effects on testosterone in rats identified by the committee's systematic review is biologically plausible. However, the mechanistic data were not sufficient to result in an upgrade in the committee's final hazard identification for fetal testosterone.

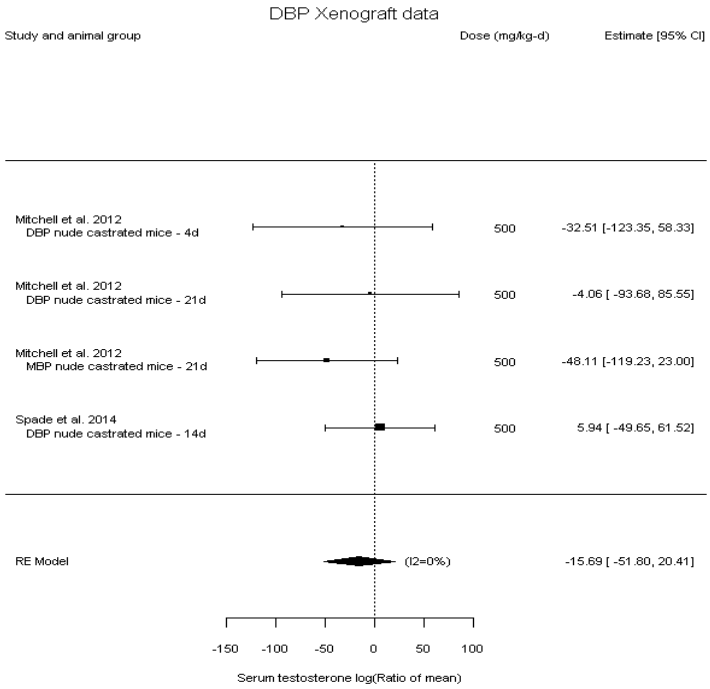


FIGURE 3-17 Meta-analysis of rodent-human xenograft studies of DBP and serum testosterone, shown as the log ratio of the mean between treated and control mice.

Final Hazard Conclusions for Fetal Testosterone

On the basis of the committee's evidence integration of the animal and the human evidence on DEHP and effects on fetal testosterone and consideration of relevant mechanistic data, the committee concluded that DEHP is presumed to be a reproductive hazard to humans.

EVIDENCE INTEGRATION FOR HYPOSPADIAS

The approach to integrating the animal and the human evidence on the effects of DEHP on fetal testosterone was the same as that used for AGD.

Initial Hazard Conclusion for Hypospadias

As described in earlier sections, the level of evidence for DEHP being associated with increased incidence of hypospadias was moderate for the animal evidence and inadequate for the human evidence. Using the OHAT hazard identification scheme (see Figure 3-4), an initial hazard conclusion is reached that DEHP is suspected to be a reproductive hazard to humans.

Consideration of Mechanistic Data

As mentioned earlier, mechanistic data available from the rat support the hypothesis that hypospadias and other phenotypic changes observed in "testicular dysgenesis syndrome" are dependent on reduction in testosterone production by the fetal Leydig cell as a proximate cause (Howdeshell et al. 2015). The results of the committee's systematic review demonstrating that DEHP induced decreases in fetal testosterone following in utero exposure provide indirect evidence for this hypothesis.

The linkage between phthalate exposure, decreased testosterone, and phenotypic changes in other species remains uncertain. For example, testicular histopathology (multinucleated gonocytes) in fetal mice exposed to DBP at 500-1,500 mg/kg-day occur in the absence of a significant decrease in testicular testosterone (Gaido et al. 2007; Lehraiki et al. 2009). Presumably, the development of microscopic evidence of urethral changes in mice (Liu et al. 2008) following DEHP exposure would also occur in the absence of decreased fetal testis testosterone production. Understanding of DEHP's effects on human fetal testosterone production and hypospadias is limited to in vitro studies and human-rodent xenograft studies that do not recapitulate the intact organism. As mentioned earlier in the discussion of the fetal testosterone data, human fetal testes cultured in vitro or implanted in rodent hosts in xenograft experiments often behave similarly to mouse testes and do not develop significant changes in fetal testosterone production following DEHP or MEHP exposure (Lambrot et al. 2009; Mitchell et al. 2012; Spade et al. 2014).

Biological Plausibility

The mechanistic data developed in vitro and in animal models provide additional evidence that the DEHP effects on testosterone and hypospadias in rats identified by the committee's systematic review are biologically plausible. Mechanistic data on DEHP effects on human fetal testosterone production and hypospadias were largely lacking, however. When considered collectively, the mechanistic data were not sufficient to result in an upgrade in the committee's final hazard identification for hypospadias.

Final Hazard Conclusions for Hypospadias

On the basis of the committee's integration of the animal evidence and the human evidence on DEHP and fetal hypospadias and consideration of relevant mechanistic data, the committee concluded that DEHP is suspected to be a reproductive hazard to humans.

CONSIDERATION OF LOW-DOSE EFFECTS

As described above, DEHP is presumed to be a reproductive hazard to humans on the basis of evaluations of the evidence on AGD and fetal testosterone. The data on AGD in animals and humans were sufficiently robust to allow quantitative characterization of the dose-response relationships, whereas the fetal testosterone evidence was not. The human data provide moderate evidence of a relationship between DEHP exposure and decreases in AGD, with a magnitude decrease of 1.7-6.5% for each $\log_{10}$ increase in exposure. When the experimental animal data were analyzed in the same manner (estimating the magnitude of change for each $\log_{10}$ increase in exposure), the effect estimates are similar: 0-2% for Sprague-Dawley rats, 1-5% for Wistar rats, and 1-3% for mice. Thus, these estimates are by and large concordant; however, the dose ranges in which these estimates have been observed differ substantially between humans and rats.

A direct comparison of DEHP exposure in animal and human studies included in the systematic reviews was not possible because exposure is measured differently in the studies. The epidemiologic studies measured DEHP metabolites in maternal urine, and none performed dose reconstruction to estimate phthalate intake or to estimate levels of active DEHP metabolites in blood or fetal testis. In contrast, animal studies reported external (administered) doses and reported no metabolite measurements in urine or other specimens. Comparison of rat administered doses and predicted human DEHP intake was, therefore, based on other studies.

Estimates of human daily intake of DEHP are generally made using either pharmacokinetic models that predict intake from urinary concentrations of phthalate metabolites ("reverse toxicokinetics") or by estimating the fraction of DEHP excreted in the urine within 24 h (Koch 2004; Lorber et al. 2010; Anderson et al. 2011; Kessler et al. 2012). Human intake estimates were available from studies that used urinary measurements reported for the general US population (Lorber et al. 2010), German subjects (Wittassek and Angerer 2008), and the Taiwanese population exposed to phthalates because of its illegal use in food

and beverages (Chang et al. 2017). Estimates of mean daily intake in adults in the US population range from approximately 0.0006-0.002 mg/kg-day (Lorber et al. 2010) to 0.011 mg/kg-day (Lorber and Calafat 2012). The study of German subjects estimated a median (maximal) daily intake of 0.003 (0.042) mg/kg-day (Wittassek and Angerer 2008). Chang et al. (2017) estimated median daily intakes of DEHP in the Taiwanese population to be about 0.004 mg/kg-day for men and 0.002 mg/kg-day for women (maximal intakes were greater than 0.008 mg/kg-day). Estimates of daily DEHP intake in all three reports were several thousand times lower than the BMD_5s of 15 and 150 mg/kg-day calculated by the committee (see Table 3-11).

Urinary MEHP concentrations in rats and humans were evaluated. Urinary MEHP concentration has been measured in rats treated with DEHP at 11 mg/kg-day (Calafat et al. 2006), a dose that is close to the lowest BMD_5 of 15 mg/kg-day calculated by the committee. Urinary MEHP concentration in pregnant rats 6 h after dosing averaged 1,626 ng/mL (Calafat et al. 2006). In humans, maternal urinary MEHP concentration was about 16.5 ng/mL (highest concentration reported for the 75th percentile group in the epidemiology studies in Table 3-11) or approximately 100 times lower than mean urinary concentrations measured in rats. Also, the 95th percentile for urinary MEHP concentrations in the general adult US population (38.9 ng/mL) is only 40 times lower than that of rats dosed at 11 mg/kg. Lorber et al. (2010) has suggested that the survey data used may underestimate US population exposure to DEHP because measurements are made in spot urine samples collected during the day, often from fasting participants. Spot urine samples have high temporal variability, especially in response to bolus dosing, so they might not provide an accurate reflection of the circulating metabolites.

A comparison of MEHP in amniotic fluid in rats dosed at 11 mg/kg-day, near the BMD_5 for decreased testosterone, and humans found that MEHP concentrations were similar. One study included in the human systematic review reported a median amniotic fluid MEHP concentration of 22.8 ng/mL (Huang et al. 2009). This concentration is about three times lower than the mean MEHP concentration (68 ng/mL) found in amniotic fluid of pregnant rats exposed at 11 mg/kg-day (Calafat et al. 2006). Slightly larger differences were found in comparisons with other human studies that were not included in the committee's systematic review (Silva et al. 2004; Wittassek et al. 2009; Huang et al. 2016).

These comparisons, qualitatively similar to margin-of-exposure comparisons, produce strikingly different results depending on whether the comparisons are based on estimates of DEHP intakes or measurement of urinary or amniotic MEHP concentration. The finding that MEHP levels in amniotic fluid in the general US population are in the same range as in rats treated near the BMD_5 for decreased fetal testosterone suggests that the dose-response relationships that emerged from the meta-analyses of the animal and human data are generally concordant. Much less is known about MEHP levels in serum or testis in humans compared with rats. As noted earlier, a striking observation in the human pharmacokinetic study is that the peak concentrations and area under the curves for MEHP and DEHP in human serum are much greater than those reported for either rats or marmosets given comparable administered doses (Koch 2004; Kessler et al. 2012). On the basis of these pharmacokinetic differences in serum MEHP and DEHP concentrations, humans could be 2-100 times more sensitive than rats to DEHP (Koch 2004; Kessler et al. 2012).

Inferences about effect levels in humans and rats are uncertain for a number of reasons. In general, the relationship between intake of DEHP and urinary levels of it and its metabolites is fairly well described, but there is much less confidence in information about blood or fetal testis concentrations in rodents or humans and their relationships with intake levels. The maternal urinary and amniotic fluid measures in rats were from a single study (Calafat et al. 2006), and the study had several limitations. Only two rats per dose group were used in the study; however, five urinary samples were collected per rat and a correlation between administered dose and urinary or amniotic fluid MEHP was found. The authors noted that MEHP was primarily found in the urine as the glucuronide conjugate, which was not consistent with other studies that reported that rats excrete free MEHP. Amniotic fluid should reflect a more integrated exposure with less variability, but there are few human studies that measure phthalate metabolites in amniotic fluid and even fewer that measured them in rats. Even though pharmacokinetic differences between humans and rats or marmosets have been demonstrated, and may explain why humans might be more

sensitive to DEHP, little information on partitioning of DEHP or its metabolites to the fetal compartment in humans is available.

OTHER PHTHALATES

This section summarizes the committee's evaluation of the other phthalates: BzBP, DBP, DEP, DIBP, DINP, and DPP. The same methods that were used to evaluate DEHP were used to evaluate these phthalates as well. Details about the risk of bias evaluations and how confidence in the body of evidence was rated are provided in Appendix C, Section C-4, for the animal studies and in Appendix D, Section D-3, for the human studies. The sections below will briefly discuss how the available data compare with DEHP and will provide initial hazard identification conclusions for AGD, fetal testosterone, and hypospadias.

Animal Health Effect Results for Other Phthalates

AGD

As with DEHP, phthalate exposure in all the animal studies of other phthalates encompassed the entirety of the male programming window, and AGD measurements taken at the earliest postnatal age were used in the analysis. For any meta-analyses performed, AGD data corrected for body weight were used whenever possible. The same risk of bias factors as with DEHP were considered, with the key factors being reliability of the outcome measure, blinding of investigators to the treatment groups, and control for litter effects. The risk of bias 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. There was no evidence of publication bias for any of the other phthalates (see Appendix C, Section C-3).

The initial rating for the confidence in the evidence from animal studies was high because they had controlled exposures, exposures occurred prior to outcome, outcomes were measured on individual animals, and a concurrent control comparison group was used (see Figure 3-2 for OHAT method for rating confidence).

BzBP and AGD

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to BzBP and effects on AGD in rodents. Six studies examining BzBP and AGD in rats were available (see Table 3-12). Five studies in rats found decreased AGD after developmental exposure to BzBP (Ashby et al. 1997; Nagao et al. 2000; Ema and Miyawaki 2002; Tyl et al. 2004; Aso et al. 2005). Confidence in the evidence was determined by considering factors that might upgrade or downgrade confidence (see Figure 3-2 for the factors). Confidence was downgraded because of risk of bias concerns; all studies had ratings of probably high risk or definitely high risk of bias in at least one of the key issues considered, and they had multiple risk of bias issues (see Figure 3-18). Confidence in the BzBP evidence was upgraded because three studies showed a relatively large magnitude of change (about 20-40%) in the same dose range, and most studies reflected a similar magnitude of response within the same dose range (see Appendix C, Figures C4-12 and C4-13).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that BzBP is associated with a decrease in AGD in male rats.

TABLE 3-11 Comparison of Human and Rat Intake and Internal Concentrations of DEHP

Human Intake or Blood Concentrations of DEHP		Rat Administered Dose, Urinary MEHP, and BMD_5	
Description (Reference)	Value	Description	Value
Intake Estimates of DEHP			
DEHP intake calculated from NHANES 2001-2002 urinary levels of MEHP, mid-range (Lorber et al. 2010)	0.0006-0.0022 mg/kg-day (mean range)	BMD_5 for effects on AGD (see Appendix C, Table C5-3)	150 mg/kg-day
DEHP intake calculated from urinary levels of MEHP in 8 adults (Lorber and Calafat 2012)	0.011 mg/kg-day (mean) 0.005 mg/kg-day (median) 0.08 mg/kg-day (max)	BMD_5 for effects on fetal testosterone (see Appendix C, Table C5-8)	15 mg/kg-day
DEHP intake calculated from urinary levels in 102 German subjects (Wittassek and Angerer 2008)	0.003 mg/kg-day (median) 0.042 mg/kg-day (max)		
DEHP intake calculated from urinary levels in a Taiwanese population (Chang et al. 2017)	Ages 7-12 years 0.005 mg/kg-day (median, males) 0.003 mg/kg-day (median, females) 0.023 mg/kg-day (95th percentile, males) 0.013 mg/kg-day (95th percentile, females) Ages 18-40 years 0.004 mg/kg-day (median, males) 0.002 mg/kg-day (median, females) 0.012 mg/kg-day (95th percentile, males) 0.018 mg//kg-day (95th percentile, females)		
Urinary Levels of MEHP			
Urinary MEHP from NHANES 2001-2002 (Lorber et al. 2010)	4.1 ng/mL (median) 38.9 ng/mL (95th percentile)	MEHP in maternal urine 6 h after exposure at 11 mg/kg-day (Calafat et al. 2006)	1,626 ng/mL (mean)
Urinary MEHP in a Taiwanese population (Huang et al. 2015)	Adults 3.4 ng/mL (GM) ND-30.5 ng/mL (5th-95th percentile) Children 4.1 ng/mL (GM) ND-27.5 ng/mL (5th-95th percentile)	MEHP in maternal urine 6 h after exposure at 100 mg/kg-day (Calafat et al. 2006)	8,107 ng/mL (mean)
Maternal urinary MEHP in Taiwanese women (Huang et al. 2016)	ND-19.8 ng/mL (25th-75th percentile)		
Maternal Urinary Levels of MEHP in Studies Included in the Systematic Review of Human Studies			
Bornehag et al. 2015	3.27 ng/mL (GM) 1.91-5.86 ng/mL (25th-75th percentile)		
Bustamante-Montes et al. 2013	4 ng/mL (mean) 0.4-19.7 ng/mL (range)		
Chevrier et al. 2012	7-19 ng/mL (33rd-66th percentile)		
Huang et al. 2009	24.9 ng/mL (median) Female fetuses: 24.6 ng/mL (median) 11.8-68.6 ng/mL (10th-90th percentile)		

	Male fetuses: 26.3 ng/mL (median) 11.9-120.3 ng/mL (10th-90th percentile)		
Jensen et al. 2016	2.0 ng/mL (mean) 0.4-2.3 ng/mL (25th-75th percentile)		
Lin et al. 2011	8.85-16.5 ng/mL (25th-75th percentile) 24.6 (median, 1st trimester) 20.6 ng/mL (median, 2nd trimester)		
Suzuki et al. 2012	3.71 ng/mL (median) 2.08-7.14 ng/mL (25th-75th percentile)		
Swan 2008	12.3 ng/mL (mean)* 6.2 ng/mL (median)*		
Swan et al. 2015; Martino-Andrade et al. 2016	First trimester: 1.93 ng/mL (GM) 4.70 ng/mL (75th percentile) Second trimester: 1.41 ng/mL (GM) Third trimester: 1.33 ng/mL (GM)		
Amniotic Fluid Levels of MEHP			
MEHP in amniotic fluid from 54 women undergoing routine amniocentesis (Silva et al. 2004)	<LOD-2.8 ng/mL (range) <LOD (10th percentile) 2.6 ng/L (95th percentile)	MEHP in rodent amniotic fluid 24 h after exposure at 11 mg/kg-day (Calafat et al. 2006)	68 ng/mL (mean)
MEHP in amniotic fluid of German women (Wittassek et al. 2009)	1.60 ng/mL (median)	MEHP in rodent amniotic fluid 24 h after exposure at 100 mg/kg-day (Calafat et al. 2006)	766 ng/mL (mean)
MEHP in 97 Taiwanese women (Huang et al. 2016)	9.62 ng/mL (75th percentile) 29.78 ng/mL (maximum)		
Amniotic Fluid Levels of MEHP in Studies Included in the Systematic Review of Human Studies			
Huang et al. 2009	22.8 ng/mL (median) Female fetuses: 24.0 ng/mL (median) 5.0-91.1 ng/mL (10th-90th percentile) Male fetuses: 22.1 ng/mL (median) 2.6-100.6 ng/mL (10th-90th percentile)		

*For boys with shorter AGD.

NOTES: CI, confidence interval; GM, geometric mean; LOD, limit of detection; ND, not detected; NHANES, National Health and Nutrition Examination Survey.

TABLE 3-12 Studies of BzBP and AGD in Rats

Study	Species	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Ahmad et al. 2014	Albino rat	GD 14-21	PND 5 and 25	100	None
Ashby et al. 1997	AP rat	GD 1 to PND 2	PND 2	None	0.18
Aso et al. 2005	Sprague-Dawley rat (F1)	GD 0 to PND 4	PND 4	400	None
	Sprague-Dawley rat (F2)	GD 0 to PND 4	PND 4	None	100
Ema and Miyawaki 2002	Wistar rat	GD 15-17	GD 21	250	500
Nagao et al. 2000	Sprague-Dawley rat	GD 0-21	PND 0	100	500
Tyl et al. 2004	Sprague-Dawley rat (F1)	GD 0-21	PND 0	50	250
	Sprague-Dawley rat (F2)	GD 0-21	PND 0	50	250

NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

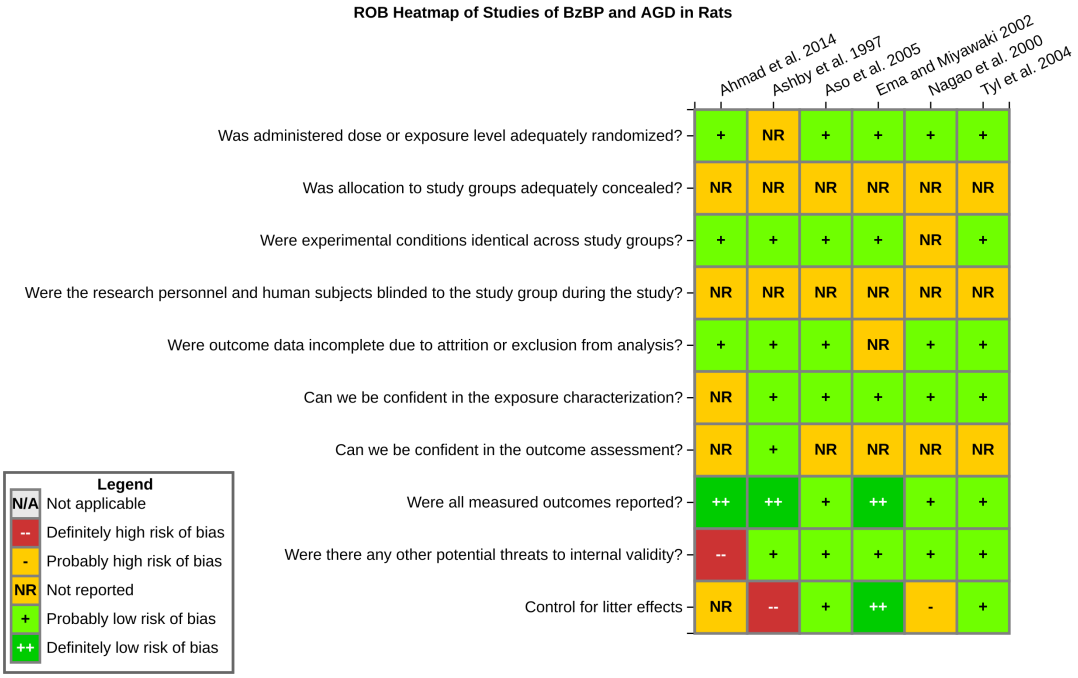


FIGURE 3-18 Risk of bias heatmap of studies of BzBP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/323/.

DBP and AGD

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DBP and effects on AGD in rodents. Twenty-two studies examining DBP and AGD in rodents were available, and multiple studies in rats found decreased AGD after developmental exposure to DBP (see Table 3-13). Confidence was downgraded because of risk of bias concerns; all studies had ratings of probably high risk or definitely high risk of bias in at least one of the key issues considered, and most of the studies had multiple risk of bias issues (see Figure 3-19). Confidence in the DBP evidence was upgraded because most studies reflected a similar magnitude of response within the same dose range (see Appendix C, Figures C4-12 and C4-13).

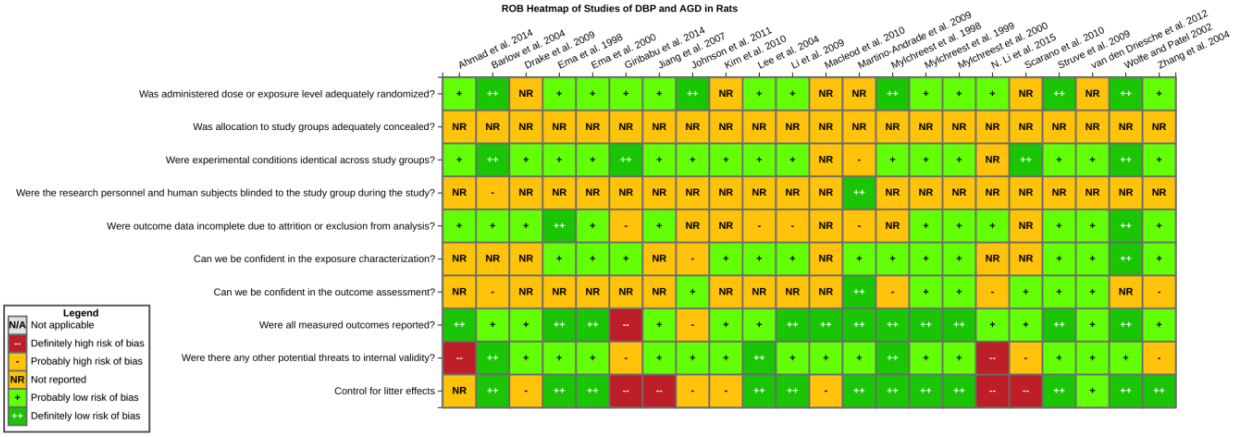


FIGURE 3-19 Risk of bias heatmap of studies of DBP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/322/.

TABLE 3-13 Studies of DBP and AGD in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Ahmad et al. 2014	Albino rat	GD 14-21	PND 5	50	None
Barlow et al. 2004	Sprague-Dawley rat	GD 12-21	PND 1	100	500
Drake et al. 2009	Wistar rat	GD 13.5-21.5	>12 weeks old	100	500
Ema et al. 1998	Wistar rat	GD 11-21	GD 21	331	555
Ema et al. 2000	Wistar rat	GD 12-14	GD 21	None	1,000
		GD 18-20	GD 21	None	1,000
		GD 15-17	GD 21	None	500
Giribabu et al. 2014	Wistar rat	GD 1, 7, and 14	PND 1	500	None
Jiang et al. 2007	Sprague-Dawley rat	GD 14-18	PND 1	250	500
Johnson et al. 2011	Fischer-344 rat	GD 12-20	GD 20	50	None
	Sprague-Dawley rat	GD 12-20	GD 20	None	500
	Sprague-Dawley rat	GD 12-20	GD 20	100	None
Kim et al. 2010	Sprague-Dawley rat	GD 10-19	PND 11	250	500
Lee et al. 2004	Sprague-Dawley rat	GD 15 to PND 2	PND 2	148	712
Li et al. 2009	Wistar rat	GD 6 to PND 1	PND 1	94	291
N. Li et al. 2015	Wistar rat	GD 12.5-20.5	PND 2	100	300
MacLeod et al. 2010	Wistar rat	GD 13.5-21.5	PND 25	100	500
Martino-Andrade et al. 2009	Wistar rat	GD 13-21	GD 21	100	500
Mylchreest et al. 1998	Sprague-Dawley rat	GD 3 to PND 1	PND 1	250	500
Mylchreest et al. 1999	Sprague-Dawley rat	GD 12-21	PND 1	100	250

(Continued)

TABLE 3-13 Continued

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Mylchreest et al. 2000	Sprague-Dawley rat	GD 12-21	PND 1	100	500
Scarano et al. 2010	Wistar rat	GD 12 to PND 4	PND 4	100	None
Struve et al. 2009	Sprague-Dawley rat	GD 12-19	GD 19	112	582
van den Driesche et al. 2012	Wistar rat	GD 19.5-20.5	GD 21.5	750	None
	Wistar rat	GD 13.5-20.5	GD 21.5	None	500
Wolfe and Patel 2002	Sprague-Dawley rat (F1a)	GD 0 to PND 1	PND 1	95	1,017
	Sprague-Dawley rat (F1b)	GD 0 to PND 1	PND 1	1,017	None
Zhang et al. 2004	Sprague-Dawley rat	GD 1 to PND 21	PND 4	50	250

NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that DBP is associated with a decrease in AGD in male rats.

DINP and AGD

Confidence in the Body of Evidence. There is very low confidence in the body of evidence on developmental exposure to DINP and effects on AGD in rodents. Four studies examining DINP and AGD in rats were available (see Table 3-14). Only one study found decreased AGD after developmental exposure to DINP (Boberg et al. 2011). Confidence was downgraded because of risk of bias concerns. Two of the studies had a probably high risk of bias rating in two key areas (whether researchers were blinded to the treatment groups or how outcomes were assessed), and one of them had a probably high risk of bias rating for not controlling for litter effects. Another study also had a no reporting about whether the researchers were blinded. Confidence in the evidence was also downgraded because of unexplained inconsistency and imprecision (see Appendix C, Figure C4-32).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a very low confidence rating in the body of evidence and questionable evidence of an effect result in a conclusion that there is an inadequate level of evidence to assess whether fetal exposure to DINP is associated with a decrease in AGD in male rats.

A summary of the confidence ratings of all the other phthalates and effects on AGD is presented in an evidence profile in Table 3-15.

TABLE 3-14 Studies of DINP and AGD in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Boberg et al. 2011	Wistar rat	GD 7 to PND 17	PND 1	750	900
Clewell et al. 2013	Sprague-Dawley rat	GD 12 to PND 14	PND 2	750	None
L. Li et al. 2015	Sprague-Dawley rat	GD 12 to 21	PND 1	1,000	None
Masutomi et al. 2003	Sprague-Dawley rat	GD 15 to PND 2	PND 2	1,165	None

NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

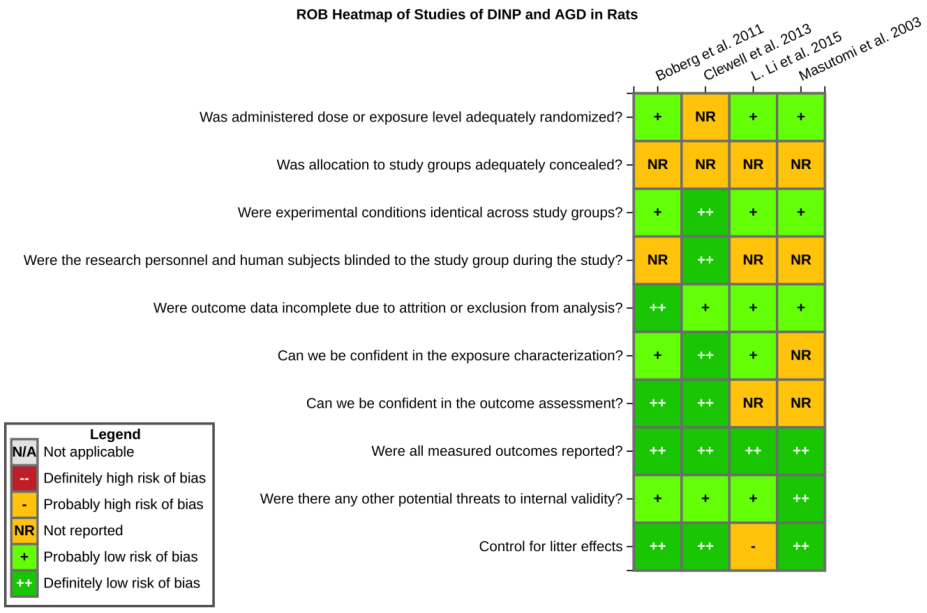


FIGURE 3-20 Risk of bias heatmap of studies of DINP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/324/.

TABLE 3-15 Profile of the Confidence in the Body of Evidence on BzBP, DBP, and DINP and AGD in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	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
BzBP	High (6 rat[a])	↓	—	—	—	—	↑	↑	—	—	—	High
DBP	High (22 rat[b])	↓	—	—	—	—	—	↑	—	—	—	High
DINP	High (4 rat[c])	↓	↓	—	↓	—	—	—	—	—	—	Very Low
DEHP	High (16 rat,[d] 3 mouse[e])	↓	—	—	—	—	↑	↑	—	—	—	High

NOTE: Evidence on DEHP included for comparison.

[a]Ashby et al. (1997); Nagao et al. (2000); Ema and Miyawaki (2002); Tyl et al. (2004); Aso et al. (2005); Ahmad et al. (2014).

[b]Ema et al. (1998, 2000); Mychreest et al. (1998, 1999, 2000); Wolfe and Patel (2002); Barlow et al. (2004); Lee et al. (2004); Zhang et al. (2004); Jiang et al. (2007): Drake et al. (2009); Li et al. (2009); Martino-Andrade et al. (2009); Struve et al. (2009); Kim et al. (2010); MacLeod et al. (2010); Scarano et al. (2010); Johnson et al. (2011); van den Driesche et al. (2012); Ahmad et al. (2014); Giribabu et al. (2014); N. Li et al. (2015).

[c]Masutomi et al. (2003); Boberg et al. (2011); Clewell et al. (2013); L. Li et al. (2015).

[d]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, 2009); Christiansen et al. (2009, 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).

[e]Liu et al. (2008); Do et al. (2012); Pocar et al. (2012).

Fetal Testosterone Concentrations

BzBP and Fetal Testosterone Concentrations

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to BzBP and effects on fetal testosterone in rats. Two studies examining BzBP and fetal testosterone in rats were available (see Table 3-16). No significant risk of bias concerns were found (both studies controlled for litter effects) or other factors that would warrant a downgrade. One study found a greater than 50% decrease in fetal testosterone in rats given BzBP at ≥100 mg/kg-day on GD 14-18 (Furr et al. 2014). Another study found a 22-90% decrease in fetal testosterone in rats given BzBP at 300-900 mg/kg-day on GD 8-18 (Howdeshell et al. 2008). Confidence in the BzBP evidence was therefore upgraded on two factors because both studies showed a relatively large magnitude of change and reflected a similar magnitude of response within the same dose range (see Appendix C, Figures C4-15 and C4-16).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to BzBP is associated with a decrease in fetal testosterone in male rats.

DBP and Fetal Testosterone Concentrations

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DBP and effects on fetal testosterone in rats. Twelve studies examining DBP and fetal testosterone in rats were available (see Table 3-17). Overall, the risk of bias concerns did not warrant a downgrade (see Figure 3-22). Multiple studies found a greater than 40% decrease in fetal testosterone in rats given DBP at ≥100 mg/kg-day during gestation. Confidence in the DBP evidence was upgraded on two factors because the studies showed a relatively large magnitude of change and reflected a similar magnitude of response within the same dose range (Appendix C, Figures C4-23 and C4-24).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DBP is associated with a decrease in fetal testosterone in male rats.

TABLE 3-16 Studies of BzBP and Fetal Testosterone in Rats

Study	Species	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Furr et al. 2014	Sprague-Dawley rat	GD 14-18	GD 18*	None	100
Howdeshell et al. 2008	Sprague-Dawley rat	GD 8-18	GD 18*	100	300

*Indicates that testes testosterone production was measured.
NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

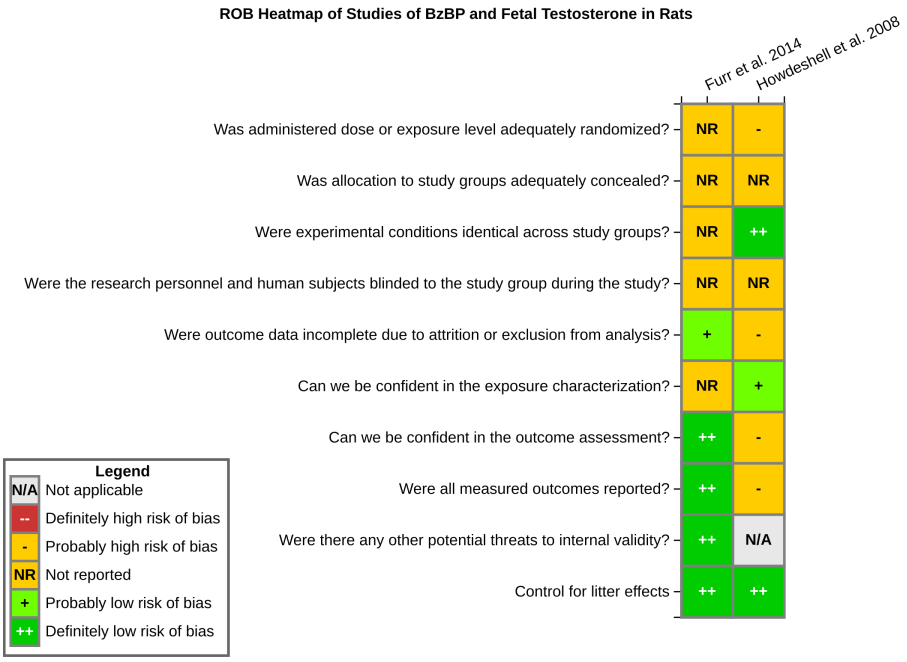


FIGURE 3-21 Risk of bias heatmap of studies of BzBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/328/.

TABLE 3-17 Studies of DBP and Fetal Testosterone in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day) [% decrease]
Clewell et al. 2009	Sprague-Dawley rat	GD 12-19	GD 19	None	100 [41]
Furr et al. 2014	Sprague-Dawley rat	GD 14-18	GD 18*	50	100 [35]
		GD 14-18	GD 18*	100	None
		GD 14-18	GD 18*	100	None
Howdeshell et al. 2008	Sprague-Dawley rat	GD 8-18	GD 18*	100	300 [34]
Johnson et al. 2007	Sprague-Dawley rat	GD 19	GD 19	100	500 [62]
Johnson et al. 2011	Fisher-344 rat	GD 12-20	GD 20	50	None
	Sprague-Dawley rat	GD 12-20	GD 20	100	None
	Sprague-Dawley rat	GD 12-20	GD 20	None	500 [85]
Kuhl et al. 2007	Sprague-Dawley rat	GD 18	GD 19	100	500 [67]
Lehmann et al. 2004	Sprague-Dawley rat	GD 12-19	GD 20	50	50 [21]
N. Li et al. 2015	Wistar rat	GD 12.5-15.5	GD 15.5	900	None
	Wistar rat	GD 12.5-17.5	GD 17.5	100	300 [78]
	Wistar rat	GD 12.5-19.5	GD 19.5	300	900 [57]
	Wistar rat	GD 12.5-21.5	GD 21.5	300	900 [57]
Mahood et al. 2007	Wistar rat	GD 13.5-20.5	GD 21.5	20	100 [14]
Martino-Andrade et al. 2009	Wistar rat	GD 13-21	GD 21	100	500 [63]
Struve et al. 2009	Sprague-Dawley rat	GD 12-19	GD 19	112	582 [96]
van den Driesche et al. 2012	Wistar rat	GD 19.5-20.5	GD 21.5	None	500 [87]

*Indicates that testes testosterone production was measured.
NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

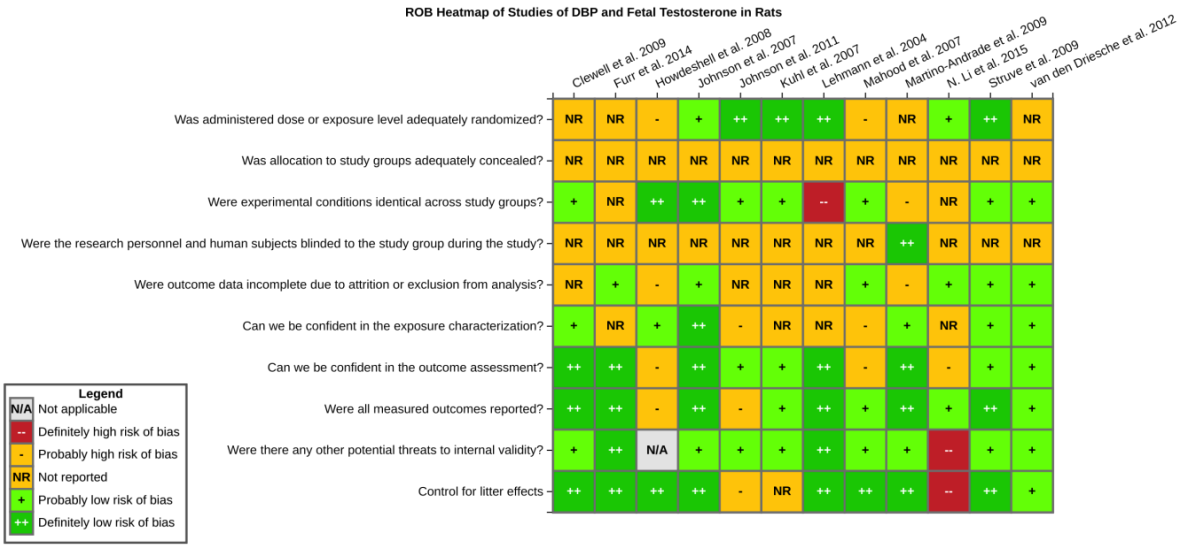


FIGURE 3-22 Risk of bias heatmap of studies of DBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/329/.

DIBP and Fetal Testosterone Concentrations

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DIBP and effects on fetal testosterone in rats. Two studies examining DBP and fetal testosterone in rats were available (see Table 3-18). Confidence in the DIBP evidence was not downgraded because of risk of bias concerns (see Figure 3-23). Confidence was downgraded because of imprecision detected in a meta-analysis of the studies (see discussion of the meta-analysis later in the chapter). Both studies found a decrease in fetal testosterone of 40% or more in rats given DIBP at ≥100 mg/kg-day during gestation (Howdeshell et al. 2008; Hannas et al. 2011b). Confidence in the DIBP evidence was upgraded for two factors because these studies showed a relatively large magnitude of change and reflected a similar magnitude of response within the same dose range (see Appendix C, Figures C4-29 and C4-30).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DIBP is associated with a decrease in fetal testosterone in male rats.

TABLE 3-18 Studies of DIBP and Fetal Testosterone in Rats

Study	Animal Group	Life Sage Exposed	Observation Time (incubation time)	NOAEL (mg/kg-day)	LOAEL (mg/kg-day) [% decrease]
Hannas et al. 2011b	Sprague-Dawley rat	GD 14-18	GD 18*	100	300 [56]
Howdeshell et al. 2008	Sprague-Dawley rat	GD 8-18	GD 18*	100	300 [40]

*Indicates that testes testosterone production was measured.
NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

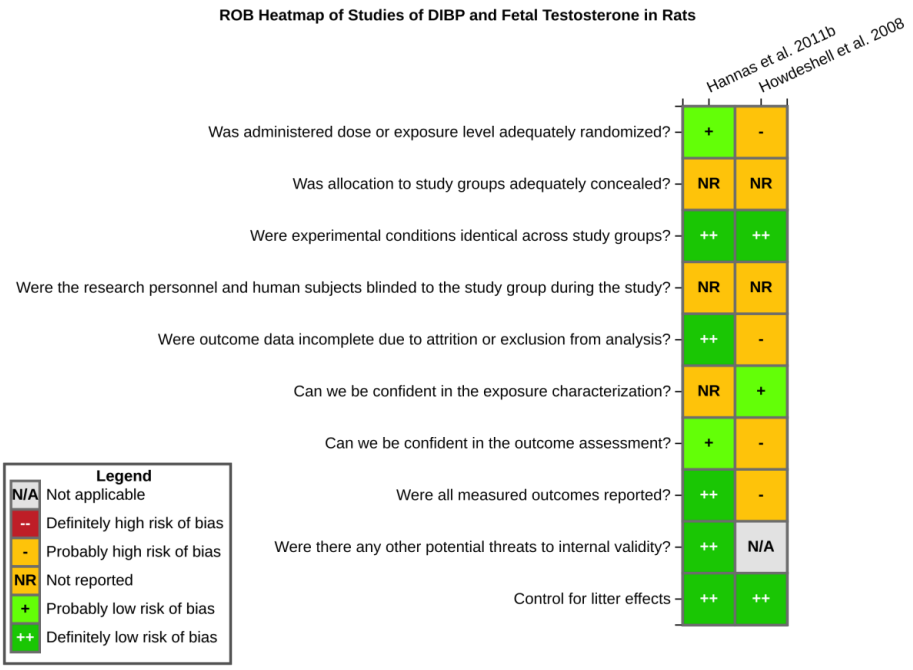


FIGURE 3-23 Risk of bias heatmap of studies of DIBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/332/.

DINP and Fetal Testosterone Concentrations

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DINP and effects on fetal testosterone in rats. Four studies examining DINP and fetal testosterone in rats were available (see Table 3-19). Confidence in the evidence was not downgraded because of risk of bias concerns (see Figure 3-24). Confidence was downgraded because of imprecision detected in a meta-analysis of the studies (see discussion of the meta-analysis later in the chapter). Confidence in the DINP evidence was upgraded on two factors because these studies reflected a large magnitude of effect and a similar magnitude of response within the same dose range (see Appendix C, Figures C4-29 and C4-30).

TABLE 3-19 Studies of DINP and Fetal Testosterone in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day) [% decrease]
Adamsson et al. 2009	Sprague-Dawley rat	GD 13.5-19.5	GD 19.5	250	750 [16]
Boberg et al. 2011	Wistar rat	GD 7-21	GD 21	300	600 [49]
Hannas et al. 2011b	Sprague-Dawley rat	GD 14-18	GD 18*	None	500 [30]
L. Li et al. 2015	Sprague-Dawley rat	GD 12-21	PND 1	500	1,000 [57]

*Indicates that testes testosterone production was measured.
NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

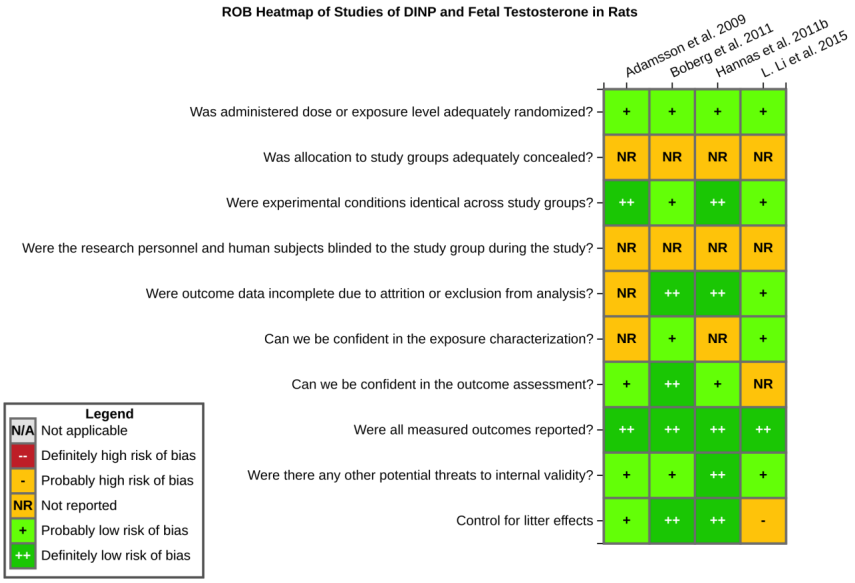


FIGURE 3-24 Risk of bias heatmap of studies of DINP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/333/.

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DINP is associated with a decrease in fetal testosterone in male rats.

DPP and Fetal Testosterone Concentrations

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DPP and effects on fetal testosterone in rats. Four studies examining DPP and fetal testosterone in rats were available (see Table 3-20). Confidence in the evidence was not downgraded because of risk of bias concerns (see Figure 3-25). Confidence in the evidence was upgraded on two factors because these studies showed a relatively large magnitude of change and reflected a similar magnitude of response within the same dose range (see Appendix C, Figures C4-38 and C4-39).

TABLE 3-20 Studies of DPP and Fetal Testosterone in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day) [% decrease]
Beverly et al. 2014	Sprague-Dawley rat	GD 14-18	GD 19*	None	50 [61]
Furr et al. 2014	Sprague-Dawley rat	GD 14-18	GD 18*	11	33 [28-62]**
		GD 14-18	GD 18*	100	None
		GD 14-18	GD 18*	100	None
Hannas et al. 2011a	Sprague-Dawley rat	GD 14-18	GD 18*	11	33 [35]
		GD 17	GD 17.5*	None	300 [29]
Howdeshell et al. 2008	Sprague-Dawley rat	GD 8-18	GD 18*	50	100 [45]

*Indicates that testes testosterone production was measured.
**Multiple blocks in this experiment.
NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

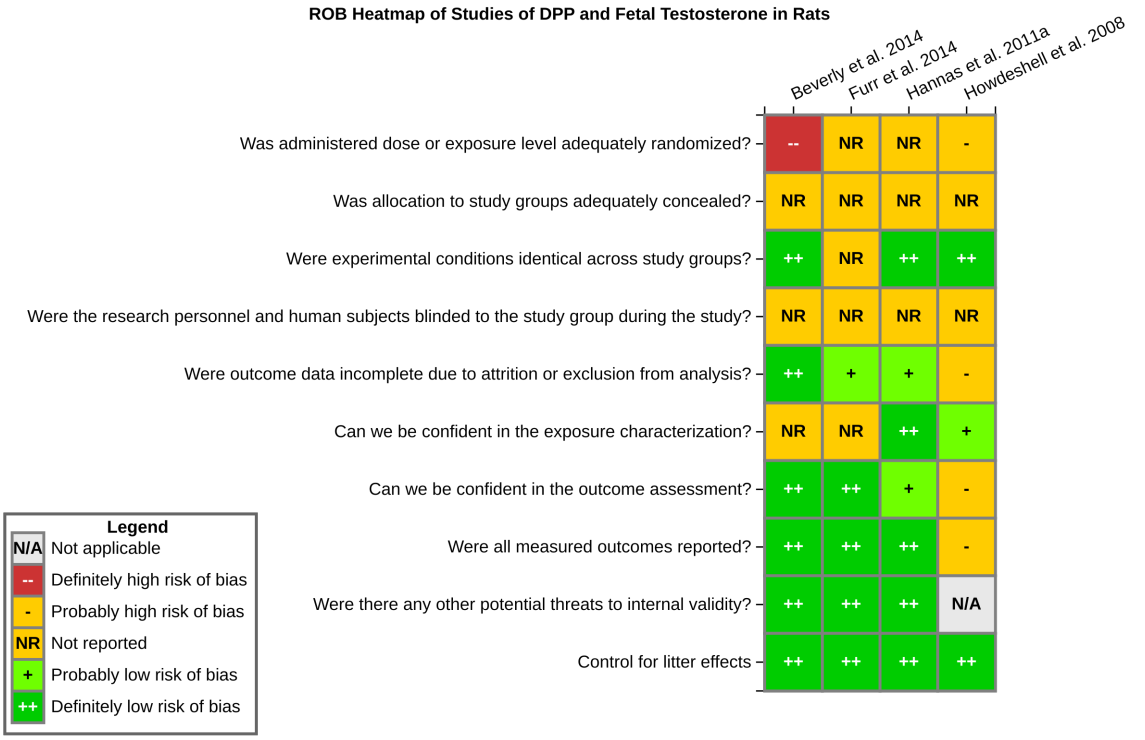


FIGURE 3-25 Risk of bias heatmap of studies of DPP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/334/.

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DPP is associated with a decrease in fetal testosterone in male rats.

A summary of the confidence ratings on all the phthalates and effects on fetal testosterone in animals is presented in an evidence profile in Table 3-21.

BzBP and Hypospadias

Confidence in the Body of Evidence. There is moderate confidence in the body of evidence on BzBP and hypospadias. Two studies examining BzBP and hypospadias in rats were found (see Table 3-22), one of which reported an increased (but not statistically significant) incidence. Confidence in the evidence was downgraded because of risk of bias concerns (see Figure 3-26); both studies had a probably high risk of bias rating because of concerns about whether the researchers were blinded to the treatment groups and concerns about the outcome measures, and one study did not control for litter effects. Because the data are limited and there were risk of bias concerns regarding the outcome measure, the committee did not upgrade confidence in the body of evidence for the finding of a rare effect as discussed earlier with DEHP phthalates and hypospadias.

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a moderate confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a moderate level of evidence that fetal exposure to BzBP is associated with an increase in hypospadias in rats.

TABLE 3-21 Profile of the Confidence in the Body of Evidence on BzBP, DBP, DIBP, DINP, and DPP and Fetal Testosterone Concentrations in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	Consistency Across Species /Model	Rare Outcome	FINAL CONFIDENCE RATING
BzBP	High (2 rat[a])	---	---	---	---	---	↑	↑	---	---	---	High
DBP	High (12 rat[b])	---	---	---	---	---	↑	↑	---	---	---	High
DIBP	High (2 rat[c])	---	---	---	↓	---	↑	↑	---	---	---	High
DINP	High (4 rat[d])	---	---	---	↓	---	↑	↑	---	---	---	High
DPP	High (4 rat[e])	---	---	---	---	---	↑	↑	---	---	---	High
DEHP	High (1 mouse,[f] 11 rat[g])	---	---	---	---	---	↑	↑	---	---	---	High

NOTE: Evidence on DEHP included for comparison.
[a]Howdeshell et al. (2008); Furr et al. (2014).
[b]Lehmann et al. (2004); Johnson et al. (2007, 2011); Kuhl et al. (2007); Mahood et al. (2007); Howdeshell et al. (2008); Clewell et al. (2009); Matrino-Andrade et al. (2009); Struve et al. (2009); van den Driesche et al. (2012); Furr et al. (2014); N. Li et al. (2015).
[c]Howdeshell et al. (2008); Hannas et al. (2011b).
[d]Adamson et al. (2009); Boberg et al. (2011); Hannas et al. (2011b); L. Li et al. (2015).
[e]Howdeshell et al. (2008); Hannas et al. (2011a); Beverly et al. (2014); Furr et al. (2014).
[f]Do et al. (2012).
[g]Borch et al. (2004, 2006); Culty et al. (2008); Howdeshell et al. (2008); Lin et al. (2008); Martino-Andrade et al. (2009); Vo et al. (2009); Hannas et al. (2011b); Klinefelter et al. (2012); Saillenfait et al. (2013a); Furr et al. (2014).

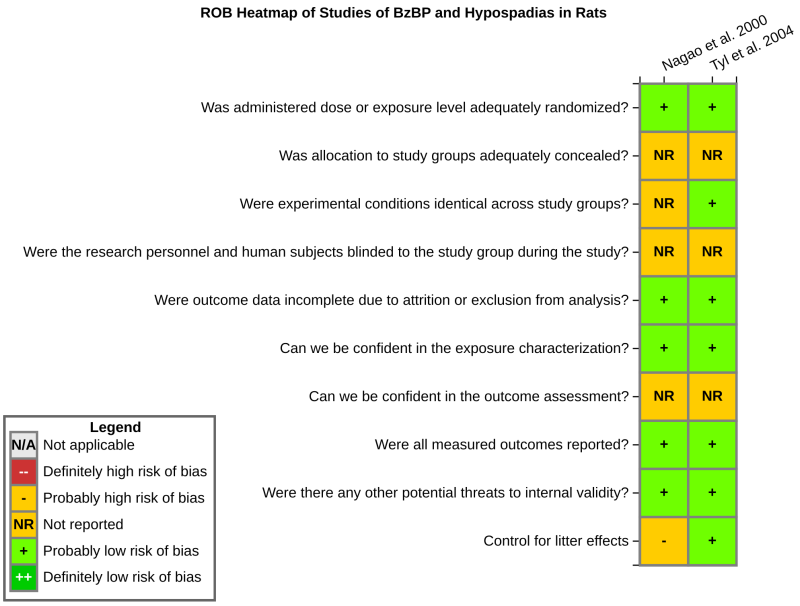


FIGURE 3-26 Risk of bias heatmap of studies of BzBP and hypospadias in rats. In HAWC: https://hawcproject.org/summary/visual/335/.

TABLE 3-22 Studies of BzBP and Hypospadias in Rats

Study	Species	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Nagao et al. 2000	Sprague-Dawley rat	GD 0-21	Postnatal	500	None
Tyl et al. 2004	Sprague-Dawley rat	GD 0-21	PND 4	750	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

DBP and Hypospadias

Confidence in the Body of Evidence. There is high confidence in the body of evidence on developmental exposure to DBP and hypospadias in rats. Eight studies examining DBP and hypospadias in rats were available (see Table 3-23). Confidence in the DBP evidence was downgraded because of risk of bias concerns (see Figure 3-27) that related to blinding of investigators and confidence in the outcome assessment. Confidence in the DPB evidence was upgraded because of a large magnitude of effect (see Appendix C, Figure C4-26) and a similar magnitude of response within the same dose range (see Appendix C, Figure C4-27). Confidence in the body of evidence was also upgraded because the background control incidence of hypospadias was reported as zero across all studies, so any positive finding was considered treatment related (i.e., rare outcome).

Level of Evidence in the Health Effect. Using the OHAT method (see Figure 3-3), a high confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a high level of evidence that fetal exposure to DBP is associated with an increase in hypospadias in rats.

A summary of the confidence ratings of all the phthalates and hypospadias in rats is presented in an evidence profile in Table 3-24.

Meta-Analyses of Animal Data

The animal database for AGD and BzBP and DBP were judged to be amenable calculated meta-analysis. Similar methods were used as previously described for DEHP. BMD_5 estimates were using a linear or linear-quadratic model, with the model selection based on the lowest AICc. The BMD_5 was calculated only for the "fixed effect"—that is, the estimated mean response across studies.

For AGD, there were statistically significant overall effects and linear trends in $\log_{10}$(dose) and dose for both BzBP and DBP. The statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. A summary of the analysis is provided in Table 3-25, and supporting details are presented in Appendix C, Section C-6.

TABLE 3-23 Studies of DBP and Hypospadias in Rats

Study	Animal Group	Life Stage Exposed	Observation Time	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Barlow et al. 2004	Sprague-Dawley rat	GD 12-21	PND 180	100	500
Drake et al. 2009	Wistar rat	GD 13.5-21.5	>12 weeks	100	500
Jiang et al. 2007	Sprague-Dawley rat	GD 14-18	PND 1	250	500
Kim et al. 2010	Sprague-Dawley rat	GD 10-19	PND 11	500	700
N. Li et al. 2015	Wistar rat	GD 12.5-20.5	PND 63	100	300
Mylchreest et al. 1998	Sprague-Dawley rat	GD 3 to PND 20	GD 100	250	500
Mylchreest et al. 1999	Sprague-Dawley rat	GD 12-21	PND 100-105	250	500
Mylchreest et al. 2000	Sprague-Dawley rat	GD 12-21	PND 110-120	100	500

*Indicates that testes testosterone production was measured.

NOTES: The earliest life stage evaluated is shown when multiple observation times were assessed. GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

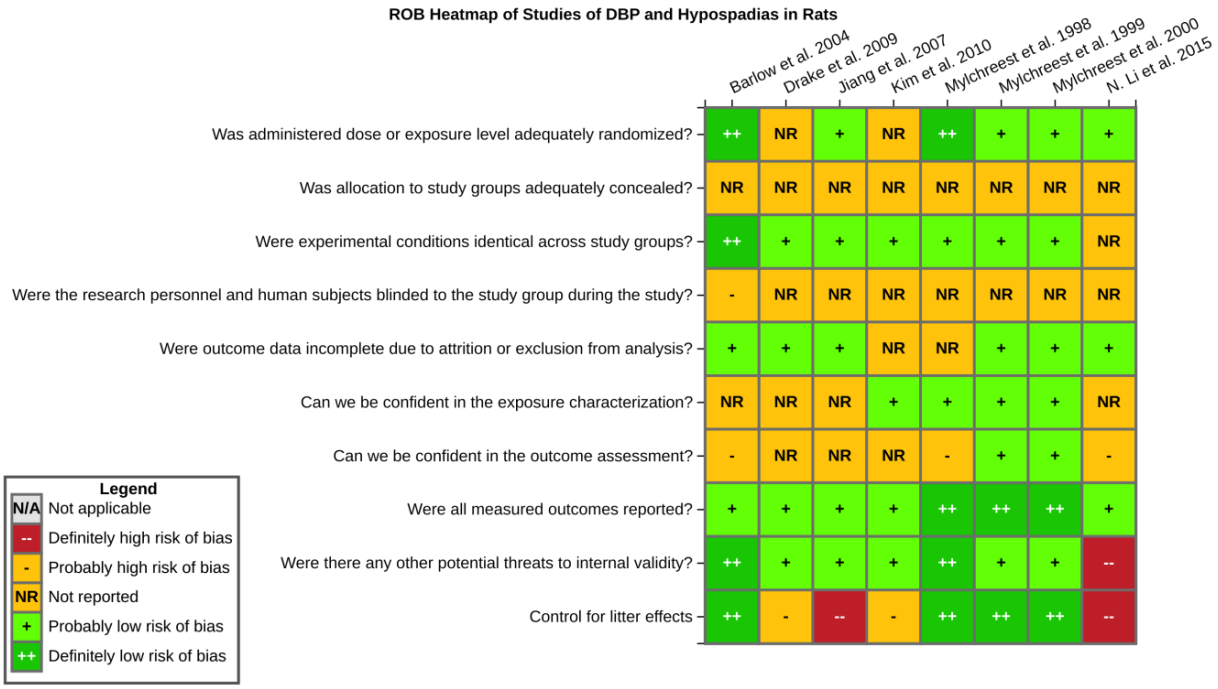


FIGURE 3-27 Risk of bias heatmap of studies on DBP and hypospadias in rats. In HAWC: https://hawcproject.org/summary/visual/338/.

TABLE 3-24 Profile of the Confidence in the Body of Evidence on BzBP and DBP and Hypospadias in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	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
BzBP	High (2 rat[a])	↓	—	—	—	—	—	—	—	—	—	Moderate
DBP	High (8 rat[b])	↓	—	—	—	—	↑	↑	—	—	↑	High
DEHP	9 (1 mouse,[c] 8 rat[d])	↓	↓	—	—	—	—	—	—	—	↑	Moderate

NOTE: Evidence on DEHP included for comparison.
[a]Nagao et al. (2000); Tyl et al. (2004).
[b]Mychreest et al. (1998, 1999, 2000); Barlow et al. (2004); Jiang et al. (2007); Drake et al. (2009); Kim et al. (2010); N. Li et al. (2015).
[c]Liu et al. (2008).
[d]Jarfelt et al. (2005); Andrade et al. (2006); Christiansen et al. (2009, 2010); Gray et al. (2009); Saillenfait et al. (2009); Vo et al. (2009); Li et al. (2013).

TABLE 3-25 Summary of Meta-Analyses for BzBP and DBP Effects on Rat AGD

Phthalate	Heterogeneity	Model with Lowest AIC	BMD_5, mg/kg-day (95% CI)
DEHP	I^2 >20%	Linear quadratic	270 (180, 420)
BzBP	I^2 >75%	Linear quadratic	250 (160, 380)
DBP	I^2 >75%	Linear quadratic	150 (120, 220)

NOTE: See Appendix C-6 for additional details. Results of the meta-analyses for DEHP (all rat strains) are included for comparison.

The animal database for fetal testosterone and BzBP, DBP, DIBP, DINP, and DPP were also judged to be amenable for meta-analysis. Similar methods were used as previously described for DEHP. BMD_5 and BMD_{40} estimates were calculated using a linear or linear-quadratic model, with the model selection based on the lowest AICc. The BMDs were calculated only for the "fixed effect"—that is, the estimated mean response across studies.

For fetal testosterone, there were statistically significant overall effects and linear trends in log10(dose) and dose for BzBP, DBP, DIBP, DINP, and DPP. The statistical significance of these effects was generally robust to leaving out individual studies and restricting to the highest dose group from each study. In the case of DIBP, there were too few studies to conduct this sensitivity analyses. A summary of the analysis is provided below (see Table 3-26), and supporting details are presented in Appendix C, Section C-6.

There were insufficient studies to perform a meta-analysis on other phthalates and hypospadias.

Human Health Effect Results on Other Phthalates

As with DEHP, the relevant human studies used state-of-the-art analytical chemistry methods to measure urinary phthalate metabolites and included collection of and adjustment for important potential confounding variables, such as measures of urinary dilution, infant body size, and maternal demographic factors. The key risk of bias evaluation factors for the human studies were whether the study designs or analyses accounted for important confounding and modifying variables, exposure characterization, and outcome assessment. There was no evidence of publication bias (see Appendix D, Table D3-1). Most studies that measured multiple urinary phthalate metabolites also measured DEHP metabolites and have been discussed previously. The initial rating for the confidence in the human studies was moderate based on the following three criteria: exposures occurred prior to outcome, outcomes were measured on individuals, and a (control) comparison group was used (see Figure 3-2 for OHAT method for rating confidence). When appropriate, meta-analyses of the human data were performed and provided additional information concerning the confidence ratings.

Reductions in AGD

Confidence in the Body of Evidence. There is moderate confidence in the body of evidence on the other phthalates and effects on AGD. Some of the prospective cohort studies of pregnant mothers and their infants that were evaluated for DEHP effects also included evaluation of association of AGD (ap or as) outcomes and other phthalate biomarkers (see Table 3-27). The relevant studies used state-of-the-art analytical chemistry methods and adjustment for important potential confounding variables. As was the case with DEHP, the effect estimates of AGD are estimates of slopes that assume a monotonic dose-response relationship between exposure and effect. Table 3-27 presents the level of confidence in the evidence for phthalate metabolites (MBP, MBzP, MCNP, MCOP, MCPP, MEP, MIBP, and MMP) and AGD in humans. The committee found no reason to upgrade or downgrade confidence in the evidence (see Appendix D, Section D-3 for details). No significant risk of bias concerns were found (see Figure 3-28).

TABLE 3-26 Summary of the Meta-Analyses for BzBP, DBP, DIBP, DINP, and DPP Effects on Rat Fetal Testosterone

Phthalate	Heterogeneity	Model with Lowest AIC	BMD_5, mg/kg-day (95% CI)	BMD_{40} mg/kg-day (95% CI)
DEHP	I^2 >90%	Linear quadratic	15 (11, 24)	160 (120, 240)
BzBP	I^2 >85%	Linear quadratic	23 (13, 74)	230 (140, 390)
DBP	I^2 >80%	Linear quadratic	12 (8, 22)	130 (85, 210)
DIBP	I^2 >60%	Linear	ND*	270 (225, 340)
DINP	I^2 >20%	Linear quadratic	76 (49, 145)	701 (552, 847)
DPP	I^2 >90%	Linear quadratic	5.6 (4.8, 6.4)	58 (50, 70)

*The 5% change was well below the range of the data, but it will be 10 times lower because a linear model was used.
NOTE: See Appendix C-6 for additional details. Results of the meta-analyses for DEHP (all rat strains) are included for comparison.

TABLE 3-27 Profile of the Confidence in the Body of Evidence on Phthalates and AGD in Humans

			Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence			
Phthalate	Metabolite(s)	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained inconsistency	Indirectness	Imprecision	Publication Bias	Large magnitude	Dose response	Residual Confounding	FINAL CONFIDENCE RATING
BzBP	MBzP	Moderate (4 prospective)[a]	—	—	—	—	—	—	—	—	Moderate
DBP	MBP	Moderate (4 prospective)[a]	—	—	—	—	—	—	—	—	Moderate
DEHP	MEHP; 5-oxo-MEHP; 5OH-MEHP; sumDEHP metabolites	Moderate (6 prospective)[b]	—	—	—	—	—	—	—	—	Moderate
DEP	MEP	Moderate (4 prospective)[a]	—	—	—	—	—	—	—	—	Moderate
DIBP	MIBP	Moderate (3 prospective)[c]	—	—	—	—	—	—	—	—	Moderate
DIDP	MCNP	Moderate (1 prospective)[d]	—	—	—	—	—	—	—	—	Moderate
DINP	MCOP	Moderate (3 prospective)[e]	—	—	—	—	—	—	—	—	Moderate
DMP	MMP	Moderate (1 prospective)[f]	—	—	—	—	—	—	—	—	Moderate
DOP	MCPP	Moderate (2 prospective)[g]	—	—	—	—	—	—	—	—	Moderate

[a]Swan (2008); Bornehag et al. (2015); Swan et al. (2015); and Jensen et al. (2016).
[b]Swan et al. (2008); Bustamante-Montes et al. (2013); Bornehag et al. (2015); Swan et al. (2015); Jensen et al. (2016); and Martino-Andrade et al. (2016).
[c]Swan (2008); Swan et al. (2015); and Jensen et al. (2016).
[d]Swan et al. (2015).
[e]Bornehag et al. (2015); Swan et al. (2015); and Jensen et al. (2016).
[f]Swan (2008).
[g]Swan (2008); Swan et al. (2015).

Meta-Analyses of Human Data on AGD and BzBP, DBP, DEP, DIBP, and DINP. Meta-analyses of human studies on BzBP, DBP, DEP, DIBP, and DINP in relation to alterations in AGD were conducted (see Appendix D, Section D-5). The same meta-analysis methods used for DEHP were applied to these phthalates. Three phthalates—DIDP, DMP, and DOP—had only one study precluding conduct of meta-analyses for these phthalates. As with the DEHP analysis, AGD (as) is preferred over AGD (ap) for each study. For the studies by Bustamonte-Montes et al. (2013) and Swan (2008), the confidence interval was estimated using the reported p-value, assuming a normal distribution. Sensitivity analyses included leaving one study out at a time and using AGD (ap) exclusively as the outcome measure. Beta coefficients standardized to a percent change per $\log_{10}$ change in metabolite exposure were used. A summary of these meta-analyses and an interpretation of the results are provided in Table 3-28.

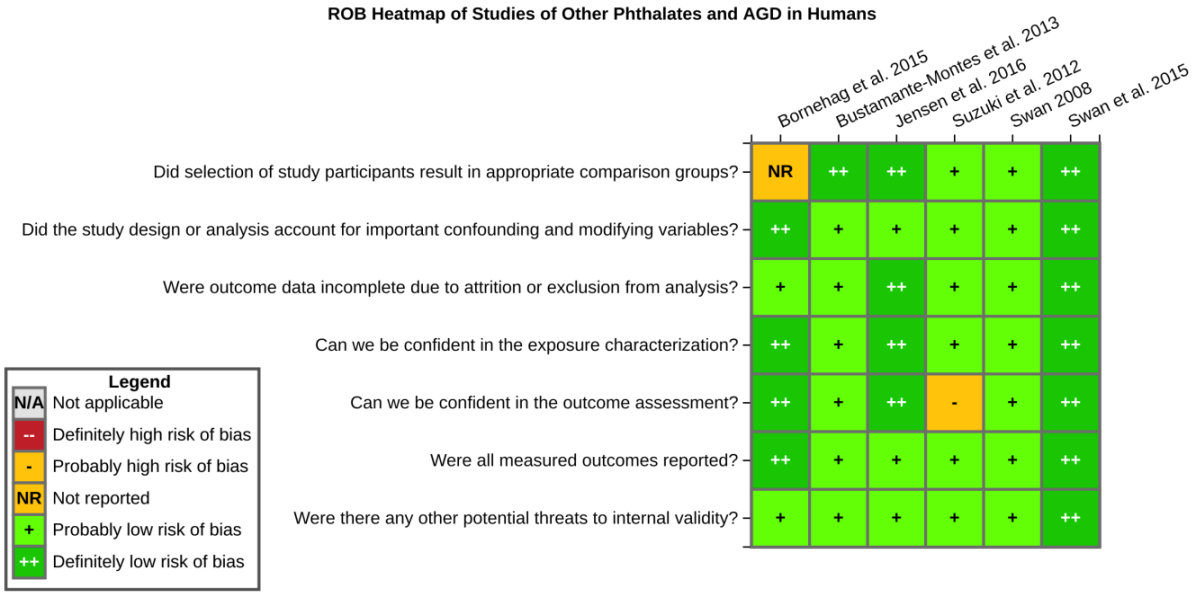


FIGURE 3-28 Risk of bias heatmap of studies of other phthalates and AGD in humans. NOTE: The study by Martino-Andrade et al. (2016) does not appear in the heatmap because it is linked to the Swan et al. (2015) study; it has the same risk of bias evaluation as that study. In HAWC: https://hawcproject.org/summary/visual/366/.

TABLE 3-28 Summary of Meta-Analyses of Human Studies of BzBP, DBP, DEP, DIBP, DINP and AGD

Phthalate (no. of studies)	Heterogeneity	Summary Estimate % change (95% CI)	Conclusion
BzBP (4)	$I^2 = 0\%$	–1.43 (–3.47, 0.61) (p = 0.17)	Available studies do not support BzBP exposure being associated with decreased AGD.
DBP (4)	$I^2 = 0\%$	–3.13 (–5.63, –0.64) (p = 0.014)	Consistent evidence of a small decrease in AGD being associated with increasing DBP exposure; magnitude around 3% for each $\log_{10}$ increase in DBP exposure.
DEP (4)	$I^2 = 29\%$	–1.94 (–3.88, 0.001) (p = 0.05)	The primary analysis suggests DEP exposure being associated with decreased AGD; effect size is small, the statistical significance of the result was not robust, and some heterogeneity was observed.
DIBP (3)	$I^2 = 0\%$	–2.23 (–5.15, 0.70) (p = 0.13)	The available studies do not support DIBP exposure being associated with decreased AGD.
DINP (3)	$I^2 = 58\%$	–0.96 (–4.17, 2.25) (p = 0.56)	The available studies do not support DINP exposure being associated with decreased AGD.

NOTE: Overall pooled estimate from random effects (RE) model per 10-fold increase in metabolite exposure is provided.

Level of Evidence in the Health Effect. Meta-analyses of the AGD studies on DBP and DEP found some evidence of a decrease in AGD being associated with exposure to these phthalates. The results show a consistent pattern of findings that higher maternal urinary concentrations of DBP and DEP during pregnancy are associated with a reduction in AGD in infancy. The small amount of heterogeneity observed for DEP may be due to sample size differences, AGD measurement variability, urinary metabolite concentration variability, and the potential for residual confounding. Using the OHAT method (see Figure 3-3), a moderate confidence rating in the body of evidence and evidence of an effect result in a conclusion that there is a moderate level of evidence that fetal exposure to DBP and DEP are associated with a reduction in AGD in male infants.

Meta-analyses of the AGD studies on BzBP, DIBP, and DINP do not show evidence of an effect. Thus, a moderate confidence rating in the body of evidence and no evidence of an effect results in a conclusion that there is an inadequate level of evidence to assess whether fetal exposure to these phthalates is associated with a decrease in AGD in male infants.

Fetal Testosterone Concentrations

Given the disparate matrices used to measure testosterone (amniotic fluid, maternal serum, or cord blood), the differences in timing of exposure (during pregnancy or at delivery), and the paucity of studies, the committee decided the data were insufficient to pursue further analysis of effects of other phthalate metabolites on fetal testosterone. (See the earlier discussion on DEHP for further details.) The committee concluded that there is inadequate evidence to determine whether fetal exposure to BzBP, DBP, DEP, DIBP, DIDP, DINP, DMP, and DOP is associated with a reduction in fetal testosterone in male humans.

Hypospadias

Given the disparate matrices used to measure phthalate metabolites (amniotic fluid and urine) and the paucity of studies, the committee decided the data were insufficient to pursue further analysis of effects of other phthalate metabolites on hypospadias. (See the earlier discussion on DEHP for further details.) The committee concluded that there is inadequate evidence to determine whether fetal exposure to BzBP, DBP, DEP, DIBP, DIDP, DINP, DMP, and DOP is associated with a hypospadias in humans.

Summary of Initial Hazard Evaluations for Other Phthalates

Table 3-29 provides initial hazard evaluations for other phthalates and AGD in humans based on the OHAT hazard identification scheme (see Figure 3-4).

RELEVANCE TO ANIMAL TOXICITY TESTING

The committee's systematic reviews and evidence integration illustrate the use of its iterative strategy for evaluating low-dose responses in humans. The rodent and human studies reviewed by the committee shared common outcome measures, AGD, changes in testosterone, and hypospadias. Despite this apparent similarity in outcomes, significant differences are present in the two data streams. The animal studies reviewed by the committee examined single phthalate exposures, whereas the human epidemiologic studies involved subjects exposed to multiple phthalates. These differences highlight the need for additional focused animal studies that more closely mimic human exposures. In addition, the human epidemiologic studies relied on biomarker data (e.g., analysis of DEHP metabolite concentrations in maternal urine during gestation as an estimate of internal dose), whereas the animal studies relied on DEHP external dose. The application of animal data to risk assessment would be strengthened by the inclusion of pharmacokinetic evaluations that could yield comparable biomonitoring data for animals that are collected in people. Despite these limitations, this case represents an example where current toxicity-testing paradigms can detect a hazard that is presumed to be of concern to humans but might not be accurately

predicting doses at which effects occur in humans. It also provides additional support for prior EPA decisions to include AGD measurements in regulatory toxicology testing (Chapter 2, Box 2-5).

Table 3-30 provides initial hazard evaluations for other phthalates and fetal testosterone in humans based on the OHAT hazard identification scheme (see Figure 3-4). Table 3-31 provides initial hazard evaluations for other phthalates and hypospadias in humans based on the OHAT hazard identification scheme (see Figure 3-4).

TABLE 3-29 Initial Hazard Evaluations for Other Phthalates and AGD in Humans

	Animal Studies		Human Studies		
Phthalate	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Initial Hazard Evaluations
BzBP	High	High	Moderate	Inadequate	Presumed human hazard
DBP	High	High	Moderate	Moderate	Presumed human hazard
DEP	Inadequate	Inadequate	Moderate	Moderate	Suspected human hazard
DIBP	Inadequate	Inadequate	Moderate	Inadequate	Not classifiable
DIDP	Inadequate	Inadequate	Inadequate	Inadequate	Not classifiable
DINP	Very Low	Inadequate	Moderate	Inadequate	Not classifiable

TABLE 3-30 Initial Hazard Evaluations for Other Phthalates and Fetal Testosterone in Humans

	Animal Studies		Human Studies		
Phthalate	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Initial Hazard Evaluations
BzBP	High	High	Inadequate	Inadequate	Presumed human hazard
DBP	High	High	Inadequate	Inadequate	Presumed human hazard
DEP	Inadequate	Inadequate	Inadequate	Inadequate	Not classifiable
DIBP	High	High	Inadequate	Inadequate	Presumed human hazard
DINP	High	High	Inadequate	Inadequate	Presumed human hazard
DPP	High	High	Inadequate	Inadequate	Presumed human hazard

TABLE 3-31 Initial Hazard Evaluations for Other Phthalates and Hypospadias in Humans

	Animal Studies		Human Studies		
Phthalate	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Level of Confidence in Evidence	Level of Evidence in the Health Effect	Initial Hazard Evaluations
BzBP	Moderate	Moderate	Inadequate	Inadequate	Suspected human hazard
DBP	High	High	Inadequate	Inadequate	Presumed human hazard

FINDINGS AND RECOMMENDATIONS

Systematic Reviews

- **Consistency and Transparency:** The committee found that the systematic review process was valuable because it provided a framework for identifying, selecting, and evaluating evidence in a consistent and explicit manner; maximized transparency in how the assessments were performed; and facilitated the clear presentation of the basis for scientific judgments.
- **Meta-analyses:** The committee found that the meta-analyses were valuable in summarizing data from the systematic reviews and in comparing the animal and the human evidence in a robust and consistent manner. For example, the meta-analyses of the animal and the human studies on

DEHP (and select other phthalates) and AGD, and of the animal studies on DEHP and (select other phthalates) and fetal testosterone, provided quantitative evidence that certain phthalates are associated with reductions in AGD and in fetal testosterone concentrations. The meta-analysis results not only informed the confidence ratings of the body of evidence but also allowed the committee to estimate benchmark doses on the basis of data from multiple studies.

Recommendation: Systematic reviews should include meta-analysis of the animal and the human evidence, if appropriate. The results of meta-analyses should be used to examine quantitative relationships between EACs and end points of interest, to inform the confidence ratings of the bodies of evidence, and, if possible, to estimate benchmark doses.

- **Risk of Bias Evaluations:** Information important to the evaluation of the quality of individual animal studies was often not reported, including whether the study controlled for litter effects, whether animals were randomly allocated to study groups, and whether research personnel were blinded to the study groups during the outcome assessment. Because a lack of adequate reporting could not be distinguished from failure to adhere to practices that minimize bias, failure to report practices that minimize bias often led to higher risk of bias ratings for individual studies, downgrading the overall level of confidence in the body of evidence. These types of problems could be remedied if journals required better reporting of the methods used in animal studies, especially reporting pertaining to issues that might introduce bias into the research. These requirements could build on reporting standards that have been developed by various organizations to improve transparency (e.g., the ARRIVE guidelines [Kilkenny et al. 2010]). For example, studies should be required to report whether animals were assigned to study groups using random allocation and whether researchers were blinded to the study groups during outcome assessment.

Evidence Integration

- A comparison of doses between animal and human studies was challenging and imprecise because animal studies characterized exposure as the administered dose (the amount of chemical that was fed or otherwise administered), whereas the human studies measured internal dose (a measurement of the chemical in a biological sample). There is some indication that the difference in internal dose between humans and rodents may be less than the difference in administered dose; these estimates are uncertain, however, and additional work is needed for clarification. Toxicology studies that measure internal dose metrics, especially with the same measure used in human biomonitoring, could help address this data gap.

  Recommendation: To support animal-to-human extrapolations, pharmacokinetic data should be generated and used to develop pharmacokinetic models that make it possible to infer human internal doses (not just intake) from biomonitoring data and animal internal doses from administered doses.

- For the evaluation of the effect of phthalates on testosterone concentrations, it was difficult to integrate findings because human studies relied on surrogate measures of in utero concentrations (testosterone measured in maternal urine, amniotic fluid, or cord serum), whereas most animal studies measured fetal testosterone in the testes of rodents. Targeted animal studies that evaluate the relationship between phthalate exposure and changes in testosterone concentrations in biological matrices more relevant to measures taken in human studies could help address this data gap.

Mechanistic Information

- Mechanistic data from animal and xenograft studies were available, and adverse outcome pathways have been proposed for how exposure to phthalates during the male programming window affects reproductive-tract development. The committee found that this information was useful during the scoping and problem formulation phase of the systematic review to help determine what outcomes should be the focus and how evidence could be integrated in reaching conclusions.
- The results of the meta-analyses and subsequent benchmark dose analyses of experimental animal data on two end points hypothesized to be part of an adverse outcome pathway (AGD and fetal testosterone concentrations) revealed species- and strain-specific quantitative differences that were not entirely consistent. For instance, there was some evidence of reduced AGD in mice exposed to DEHP, but without evidence of decreased testosterone (or increased hypospadias). Additionally, compared to Sprague-Dawley rats, Wistar rats appeared to be more sensitive to changes in AGD but less sensitive to decreases in testosterone and much less sensitive to hypospadias.
- Given the variation in sensitivities among rodent species and strains, the mechanistic information was less useful in considering concordance and discordance between the animal and the human studies beyond providing evidence of the biological plausibility that the effects observed in rat studies identify the same hazards in humans.
- A meta-analysis of xenograft data on DBP and serum testosterone was performed to show how meta-analyses could be applied to mechanistic studies.

Hazard Identification

- The committee concluded that exposure to DEHP and certain other phthalates are presumed to cause decreases in AGD in humans, based on a moderate level of evidence from human studies and a high level of evidence from animal studies. Measurement of AGD in rodents has value in the identification of reproductive hazards in humans associated with EACs.
- The committee concluded that exposure to DEHP and certain other phthalates are presumed to cause decreases in fetal testes testosterone, based on inadequate evidence from human studies and a high level of evidence from animal studies.
- The committee concluded that exposure to DEHP and certain other phthalates is suspected to cause increases in the risk of hypospadias in humans, based on inadequate evidence from human studies and a moderate level of evidence from animal studies.

Low-Dose Effects

- The committee concluded that the human studies provide a moderate level of evidence that fetal exposure to DEHP is associated with decreases in AGD in humans. Uncertainty in the internal doses of humans relative to experimental animals limited the ability to draw conclusions about the prediction of low-dose effects based on experimental animal studies. The development of pharmacokinetic data and models for extrapolation of data from animal studies or human biomonitoring data could facilitate the evaluation of the potential of phthalates to cause health effects in humans at low doses.

Other Considerations

- **Mixtures:** The committee found that humans are exposed to a mixture of phthalates, whereas the experimental animal evidence was generally from studies with single phthalate congeners.

This difference between human mixture exposure and the single chemical animal exposures contributed to the challenges for integrating evidence between human and animal studies.

- **Expertise:** The committee found that the conduct of the systematic review and evidence integration requires a multidisciplinary approach that should be tailored to the specific review question. Experts in the conduct of meta-analyses and benchmark dose modeling will be essential.

REFERENCES

Adachi, K., H. Suemizu, N. Murayama, M. Shimizu, and H. Yamazaki. 2015. Human biofluid concentrations of mono(2-ethylhexyl)phthalate extrapolated from pharmacokinetics in chimeric mice with humanized liver administered with di(2-ethylhexyl)phthalate and physiologically based pharmacokinetic modeling. Environ. Toxicol. Pharmacol. 39(3):1067-1073.

Adamsson, A., V. Salonen, J. Paranko, and J. Toppari. 2009. Effects of maternal exposure to di-isononylphthalate (DINP) and 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (p,p'-DDE) on steroidogenesis in the fetal rat testis and adrenal gland. Reprod. Toxicol. 28(1):66-74.

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. J. Clin. Endocrinol. Metab. 100(9):E1216-E1224.

Ahmad, R., A.K. Gautam, Y. Verma, S. Sedha, and S. Kumar. 2014. Effects of in utero di-butyl phthalate and butyl benzyl phthalate exposure on offspring development and male reproduction of rat. Environ. Sci. Pollut. Res. Int. 21(4):3156-3165.

Albro, P.W., and S.R. Lavenhar. 1989. Metabolism of di(2-ethylhexyl) phthalate. Drug Metab. Rev. 21(1):13-34.

Anderson, W.A., L. Castle, S. Hird, J. Jeffery, and M.J. Scotter. 2011. A twenty-volunteer study using deuterium labelling to determine the kinetics and fractional excretion of primary and secondary urinary metabolites of di-2-ethylhexylphthalate and di-iso-nonylphthlate. Food Chem. Toxicol. 49(9):2022-2029.

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(1):64-74.

Araki, A., T. Mitsui, C. Miyashita, T. Nakajima, H. Naito, S. Ito, S. Sasaki, K. Cho, T. Ikeno, K. Nonomura, and R. Kishi. 2014. Association between maternal exposure to di(2-ethylhexyl) phthalate and reproductive hormone levels in fetal blood: The Hokkaido study on environment and children's health. PLOS One 9(10):e109039.

Ashby, J., H. Tinwell, P.A. Lefevre, J. Odum, D. Paton, S.W. Millward, S. Tittensor, and A.N. Brooks**.** 1997. Normal sexual development of rats exposed to butyl benzyl phthalate from conception to weaning. Regul. Toxicol. Pharmacol. 26(1 Pt. 1):102-118.

Aso, S., H. Ehara, K. Miyata, S. Hosyuyama, K. Shiraishi, T. Umano, and Y. Minobe. 2005. A two-generation reproductive toxicity study of butyl benzyl phthalate in rats. J. Toxicol. Sci. 30(Spec.):39-58.

Ban, S., F. Sata, N. Kurahasi, S. Kasai, K. Moriya, H. Kakizaki, K. Nonomura, and R. Kishi. 2008. Genetic polymorphisms of ESR1 and ESR2 that may influence estrogen activity and the risk of hypospadias. Hum. Reprod. 23(6):1466-1471.

Barlow, N.J., B.S. McIntyre, and P.M. Foster. 2004. Male reproductive tract lesions at 6, 12, and 18 months of age following in utero exposure to di(n-butyl) phthalate. Toxicol. Pathol. 32(1):79-90.

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. Paediatr. Perinat. Epidemiol. 30(2):105-114.

Beverly, B.E., C.S. Lambright, J.R. Furr, H. Sampson, V.S. Wilson, B.S. McIntyre, P.M. Foster, G. Travlos, and L.E. Gray, Jr. 2014. Simvastatin and dipentyl phthalate lower ex vivo testicular testosterone production and exhibit additive effects on testicular testosterone and gene expression via distinct mechanistic pathways in the fetal rat. Toxicol. Sci. 141(2):524-537.

Boberg, J., S. Christiansen, M. Axelstad, T.S. Kledal, A.M. Vinggaard, M. Dalgaard, C. Nellemann, and U. Hass. 2011. Reproductive and behavioral effects of diisononyl phthalate (DINP) in perinatally exposed rats. Reprod. Toxicol. 31(2):200-209.

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. Reprod. Toxicol. 18(1):53-61.

Borch, J., S.B. Metzdorff, A.M. Vinggaard, L. Brokken, and M. Dalgaard. 2006. Mechanisms underlying the anti-androgenic effects of diethylhexyl phthalate in fetal rat testis. Toxicology 223(1-2):144-155.

Bornehag, C.G., F. Carlstedt, B.A. Jonsson, 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. Environ. Health Perspect. 123(1):101-107.

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. J. Dev. Orig. Health Dis. 4(4):300-306.

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(1):22-30.

Calafat, A.M., M.P. Longnecker, H.M. Koch, S.H. Swan, R. Hauser, L.R. Goldman, B.P. Lanphear, R.A. Rudel, S.M. Engel, S.L. Teitelbaum, R.M. Whyatt, and M.S. Wolff. 2015. Optimal exposure biomarkers for nonpersistent chemicals in environmental epidemiology. Environ. Health Perspect. 123(7):A166-A168.

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]. Available: http://www.cdc.gov/exposurereport/ [accessed June 15, 2016].

CDC. 2015. Fourth Report on Human Exposure to Environmental Chemicals, Updated Tables, February 2015. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention [online]. Available: http://www.cdc.gov/exposurereport/ [access June 15, 2016].

Chang, J.W., C.C. Lee, W.H. Pan, W.C. Chou, H.B. Huang, H.C. Chiang, and P.C. Huang. 2017. Estimated daily intake and cumulative risk assessment of phthalates in the general Taiwanese after the 2011 DEHP food scandal. Sci. Rep. 7:45009.

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. U.S. Consumer Product Safety Commission, Directorate for Health Sciences, Bethesda, MD [online]. Available: https://www.cpsc.gov/PageFiles/169902/CHAP-REPORT-With-Appendices.pdf [accessed June 17, 2016].

Chauvigné, F., A. Menuet, L. Lesné, M.C. Chagnon, C. Chevrier, J.F. Regnier, J. Angerer, and B. Jégou. 2009. Time- and dose-related effects of di-(2-ethylhexyl) phthalate and its main metabolites on the function of the rat fetal testis in vitro. Environ. Health Perspect. 117(4):515-521.

Chevrier, C., C. Petit, C. Philippat, M. Mortamais, R. Slama, F. Rouget, A.M. Calafat, X. Ye, M.J. Silva, M.A. Charles, and S. Cordier. 2012. Maternal urinary phthalates and phenols and male genital anomalies. Epidemiology 23(2):353-356.

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. Environ. Health Perspect. 117(12):1839-1846.

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. Reprod. Toxicol. 30(2):313-321.

Clewell, R.A., J.J. Kremer, C.C. Williams, J.L. Campbell, M.A. Sochaski, M.E. Andersen, and S.J. Borghoff. 2009. Kinetics of selected di-n-butyl phthalate metabolites and fetal testosterone following repeated and single administration in pregnant rats. Toxicology 255(1-2):80-90.

Clewell, R.A., J.L. Campbell, S.M. Ross, K.W. Gaido, H.J. Clewell, III, and M.E. Andersen. 2010. Assessing the relevance of in vitro measures of phthalate inhibition of steroidogenesis for in vivo response. Toxicol. In Vitro 24(1):327-334.

Clewell, R.A., A. Thomas, G. Willson, D.M. Creasy, and M.E. Andersen. 2013. A dose response study to assess effects after dietary administration of diisononyl phthalate (DINP) in gestation and lactation on male rat sexual development. Reprod. Toxicol. 35:70-80.

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. Biol. Reprod. 78(6):1018-1028.

Dean, A., and R.M. Sharpe. 2013. Clinical review: Anogenital distance or digit length ratio as measures of fetal androgen exposure: Relationship to male reproductive development and its disorders. J. Clin. Endocrinol. Metab. 98(6):2230-2238.

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. Reprod. Toxicol. 34(4):614-621.

Drake, A.J., S. van den Driesche, H.M. Scott, G.R. Hutchison, J.R. Seckl, and R.M. Sharpe**.** 2009. Glucocorticoids amplify dibutyl phthalate-induced disruption of testosterone production and male reproductive development. Endocrinology 150(11):5055-5064.

Ema, M., and E. Miyawaki. 2002. Effects on development of the reproductive system in male offspring of rats given butyl benzyl phthalate during late pregnancy. Reprod. Toxicol. 16(1):71-76.

Ema, M., E. Miyawaki, and K. Kawashima. 1998. Further evaluation of developmental toxicity of di-n-butyl phthalate following administration during late pregnancy in rats. Toxicol. Lett. 98(1-2):87-93.

Ema, M., E. Miyawaki, and K. Kawashima. 2000. Critical period for adverse effects on development of reproductive system in male offspring of rats given di-n-butyl phthalate during late pregnancy. Toxicol. Lett. 111(3):271-278.

EU (European Union). 2004. Commission Directive 2004/93/EC of 21 September 2004 amending Council Directive 76/768/EEC for the purpose of adapting its Annexes II and III to technical progress. O.J. EU I.300:13-41. September 25, 2004 [online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:300:0013:0041:EN:PDF [accessed February 13, 2017].

EU. 2005a. Directive 2005/84/EC of the European Parliament and of the Council 14 December 2005 amending for the 22nd time Council Directive 76/769/EEC on the approximation of the laws, regulations and administrative provisions of the member states relating to restrictions on the marketing and use of certain dangerous substances and preparations (phthalates in toys and childcare articles). O.J. EU L344:40-43. December 27, 2005 [online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:344:0040:0043:EN:PDF [accessed February 13, 2107].

EU. 2005b. Commission Directive 2005/80/EC of 21 November 2005 amending Council Directive 76/768/EEC, concerning cosmetic products, for the purposes of adapting Annexes II and III there to technical progress. O.J. EU I.303:32-37. November 22, 2005 [online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005L0080&from=EN [accessed February 13, 2017].

EU. 2007. Commission Directive 2007/19/EC of 30 March 2007 amending Directive 2002/72/EC relating to plastic materials and articles intended to come into contact with food and Council Directive 85/572/EEC laying down the list of simulants to be used for testing migration of constituents of plastic materials and articles intended to come into contact with foodstuffs. O. J. EU L91:17-36. March 31, 2007 [online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:091:0017:0036:E N:PDF [accessed February 13, 2017].

Fennell, T.R., W.L. Krol, S.C. Sumner, and R.W. Snyder. 2004. Pharmacokinetics of dibutyl phthalate in pregnant rats. Toxicol. Sci. 82(2):407-418.

Fisher, J.S., S. Macpherson, N. Marchetti, and R.M. Sharpe. 2003. Human "testicular dysgenesis syndrome": A possible model using in-utero exposure of the rat to dibutyl phthalate. Hum. Reprod. 18(7):1383-1394.

Foster, P.M., E. Mylchreest, K.W. Gaido, and M. Sar. 2001. Effects of phthalate esters on the developing reproductive tract of male rats. Hum. Reprod. Update 7(3):231-235.

Fujii, S., K. Yabe, M. Furukawa, M. Hirata, M. Kiguchi, and T. Ikka. 2005. A two-generation reproductive toxicity study of diethyl phthalate (DEP) in rats. J. Toxicol. Sci. 30(Spec.):97-116.

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. Toxicol. Sci. 140(2):403-424.

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. Toxicol. Sci. 97(2):491-503.

Giribabu, N., S.B. Sainath, and P.S. Reddy. 2014. Prenatal di-n-butyl phthalate exposure alters reproductive functions at adulthood in male rats. Environ. Toxicol. 29(5):534-544.

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. Toxicol. Sci. 58(2):350-365.

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. Toxicol. Sci. 110(2):411-425.

Gray, L.E., Jr., J. Furr, K.R. Tatum-Gibbs, C. Lambright, H. Sampson, B.R. Hannas, V.S. Wilson, A. Hotchkiss, and P.M. Foster. 2016. Establishing the "biological relevance" of dipentyl phthalate reductions in fetal rat testosterone production and plasma and testis testosterone levels. Toxicol. Sci. 149(1):178-191.

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

Hannas, B.R., J. Furr, C.S. Lambright, V.S. Wilson, P.M. Foster, and L.E. Gray, Jr. 2011a. Dipentyl phthalate dosing during sexual differentiation disrupts fetal testis function and postnatal development of the male Sprague-Dawley rat with greater relative potency than other phthalates. Toxicol. Sci. 120(1):184-193.

Hannas, B.R., C.S. Lambright, J. Furr, K.L. Howdeshell, V.S. Wilson, and L.E. Gray, Jr. 2011b. Dose-response assessment of fetal testosterone production and gene expression levels in rat testes following in utero exposure to diethylhexyl phthalate, diisobutyl phthalate, diisoheptyl phthalate, and diisononyl phthalate. Toxicol. Sci. 123(1):206-216.

Hannas, B.R., C.S. Lambright, J. Furr, N. Evans, P.M. Foster, E.L. Gray, and V.S. Wilson. 2012. Genomic biomarkers of phthalate-induced male reproductive developmental toxicity: A targeted RT-PCR array approach for defining relative potency. Toxicol. Sci. 125(2):544-557.

Hauser, R., and A.M. Calafat. 2005. Phthalates and human health. Occup. Environ. Med. 62(11):806-818.

Hedges, L.V., J. Gurevitch, and P.S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80(4):1150-1156.

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. Environ. Health Perspect. 120(80):1137-1143.

Higgins, J.P.T., and S. Green, eds. 2011. Identifying and measuring heterogeneity. Chapter 9 in Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available: http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm [accessed March 8, 2017].

Howdeshell, K.L., V.S. Wilson, J. Furr, C.R. Lambright, C.V. Rider, C.R. Blystone, A.K. Hotchkiss, and L.E. Gray, Jr. 2008. A mixture of five phthalate esters inhibits fetal testicular testosterone production in the Sprague-Dawley rat in a cumulative, dose-additive manner. Toxicol. Sci. 105(1):153-165.

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. Toxicol. Sci. 148(2):488-502.

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. Hum. Reprod. 27(6):1577-1580.

Huang, P.C., P.L. Kuo, Y.Y. Chou, S.J. Lin, and C.C. Lee. 2009. Association between prenatal exposure to phthalates and the health of newborns. Environ. Int. 35(1):14-20.

Huang, P.C., C.H. Tsai, W.Y. Liang, S.S. Li, W.H. Pan, H.C. Chiang. 2015. Age and gender differences in urinary levels of eleven phthalate metabolites in general Taiwanese population after a DEHP episode. PLoS ONE 10(7):e0133782.

Huang, P.C., C.H. Tsai, W.Y. Liang, S.S. Li, H.B. Huang, and P.L. Kuo. 2016. Early phthalates exposure in pregnant women is associated with alteration of thyroid hormones. PLoS ONE 11 (7):e0159398.

Ito, Y., H. Yokota, R. Wang, O. Yamanoshita, G. Ichihara, H. Wang, Y. Kurata, K. Takagi, and T. Nakajima. 2005. Species differences in the metabolism of di(2-ethylhexyl) phthalate (DEHP) in several organs of mice, rats, and marmosets. Arch. Toxicol. 79(3):147-154.

Ito, Y., M. Kamijima, C. Hasegawa, M. Tagawa, T. Kawai, M. Miyake, Y. Hayashi, H. Naito, and T. Nakajima. 2014. Species and inter-individual differences in metabolic capacity of di(2-ethylhexyl)phthalate (DEHP) between human and mouse livers. Environ. Health Prev. Med. 19(2):117-125.

Jain, V.G., and A.K. Singal. 2013. Shorter anogenital distance correlates with undescended testis: A detailed genital anthropometric analysis in human newborns. Hum. Reprod. 28(9):2343-2349.

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. Reprod. Toxicol. 19(4):505-515.

Jensen, M.S., R. Anand-Ivell, B. Norgaard-Pedersen, B.A. Jonsson, J.P. Bonde, D.M. Hougaard, A. Cohen, C.H. Lindh, R. Ivell, and G. Toft. 2015. Amniotic fluid phthalate levels and male fetal gonad function. Epidemiology 26(1):91-99.

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, and A.M. Andersson. 2016. Prenatal exposure to phthalates and anogenital distance in male infants from a low-exposed Danish cohort (2010-2012). Environ. Health Perspect. 124(7):1107-1113.

Jiang, J., L. Ma, L. Yuan, X. Wang, and W. Zhang. 2007. Study on developmental abnormalities in hypospadiac male rats induced by maternal exposure to di-n-butyl phthalate (DBP). Toxicology 232(3):286-293.

Johns, L.E., G.S. Cooper, A. Galizia, and J.D. Meeker. 2015. Exposure assessment issues in epidemiology studies of phthalates. Environ Int. 85:27-39.

Johnson, K.J., J.B. Hensley, M.D. Kelso, D.G. Wallace, and K.W. Gaido. 2007. Mapping gene expression changes in the fetal rat testis following acute dibutyl phthalate exposure defines a complex temporal cascade of responding cell types. Biol. Reprod. 77(6):978-989.

Johnson, K.J., E.N. McDowell, M.P. Viereck, and J.Q. Xia. 2011. Species-specific dibutyl phthalate fetal testis endocrine disruption correlates with inhibition of SREBP2-dependent gene expression pathways. Toxicol. Sci. 120(2):460-474.

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. Toxicol. Sci.129(2):235-248.

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. Biol. Reprod. 93(4):92.

Karmaus, A.L., C.M. Toole, D.L. Filer, K.C. Lewis, and M.T. Martin. 2016. High-throughput screening of chemical effects on steroidogenesis using H295R human adrenocortical carcinoma cells. Toxicol. Sci. 150(2):323-332.

Kay, V.R., M.S. Bloom, and W.G. Foster. 2014. Reproductive and developmental effects of phthalate diesters in males. Crit. Rev. Toxicol. 44(6):467-498.

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. Toxicol. Appl. Pharmacol. 195(2):142-153.

Kessler, W., W. Numtip, W. Volkel, E. Seckin, G.A. Csanady, C. Putz, D. Klein, H. Fromme, and J.G. Filser. 2012. Kinetics of di(2ethylhexyl) phthalate (DEHP) and mon(2-ethylhexyl) phthalate in blood and of DEHP metabolites in urine of male volunteers after single ingestion of ring-deuterated DEHP. Toxicol. Appl. Pharmacol. 264(2):284-291.

Kilkenny, C., W.J. Browne, I.C. Cuthill, M. Emerson, and D.G. Altman. 2010. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLOS Biol. 8(6):e1000412.

Kim, T.S., K.K. Jung, S.S. Kim, I.H. Kang, J.H. Baek, H.S. Nam, S.K. Hong, B.M. Lee, J.T. Hong, K.W. Oh, H.S. Kim, S.Y. Han, and T.S. Kang. 2010. Effects of in utero exposure to di(n-butyl) phthalate on development of male reproductive tracts in Sprague-Dawley rats. J. Toxicol. Environ. Health A 73(21-22):1544-1559.

Klinefelter, G.R., J.W. Laskey, W.M. Winnik, J.D. Suarez, N.L. Roberts, L.F. Strader, B.W. Riffle, and D.N. Veeramachaneni. 2012. Novel molecular targets associated with testicular dysgenesis induced by gestational exposure to diethylhexyl phthalate in the rat: A role for estradiol. Reproduction 144(6):747-761.

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. Arch. Toxicol. 78(3):123-130.

Kuhl, A.J., S.M. Ross, and K.W. Gaido. 2007. CCAAT/enhancer binding protein β, but not steroidogenic factor-1, modulates the phthalate-induced dysregulation of rat fetal testicular steroidogenesis. Endocrinology 148(12):5851-5864.

Kurata, Y., F. Makinodan, N. Shimamura, and M. Katoh. 2012. Metabolism of di (2-ethylhexyl) phthalate (DEHP): Comparative study in juvenile and fetal marmosets and rats. J. Toxicol. Sci. 37(1):33-49.

Lajeunesse, M.J. 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92(11):2049-2055.

Lambrot, R., V. Muczynski, C. Lécureuil, G. Angenard, H. Coffigny, C. Pairault, D. Moison, R. Frydman, R. Habert, and V. Rouiller-Fabre. 2009. Phthalates impair germ cell development in the human fetal testis in vitro without change in testosterone production. Environ. Health Perspect. 117(1):32-37.

Latini, G. 2005. Monitoring phthalate exposure in humans. Clin. Chim. Acta 361(1-2):20-29.

Lee, K.Y., M. Shibutani, H. Takagi, N. Kato, S. Takigami, C. Uneyama, and M. Hirose. 2004. Diverse developmental toxicity of di-n-butyl phthalate in both sexes of rat offspring after maternal exposure during the period from late gestation through lactation. Toxicology 203(1-3):221-238.

Lehmann, K.P., S. Phillips, M. Sar, P.M. Foster, and K.W. Gaido. 2004. Dose-dependent alterations in gene expression and testosterone synthesis in the fetal testes of male rats exposed to di(n-butyl) phthalate. Toxicol. Sci. 81(1):60-68.

Lehraiki, A., C. Racine, A. Krust, R. Habert, and C. Levacher. 2009. Phthalates impair germ cell number in the mouse fetal testis by an androgen- and estrogen-independent manner. Toxicol. Sci. 111(2):372-382.

Li, H. , Y. Li, J. Fang, K. Ma, S. Wu, R. Gao, X. Yan, J. Cheng, J. Liu, and F. Dong. 2014. Analysis of four phthalate monoesters in human urine using liquid chromatography–tandem mass spectrometry. LCGC North America 32(6):434-439.

Li, L., T. Bu, H. Su, Z. Chen, Y. Liang, G. Zhang, D. Zhu, Y. Shan, R. Xu, Y. Hu, J. Li, G. Hu, Q. Lian, and R.S. Ge. 2015. In utero exposure to diisononyl phthalate caused testicular dysgenesis of rat fetal testis. Toxicol. Lett. 232(2):466-474.

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. Environ. Toxicol. Pharmacol. 35(1):55-60.

Li, N., X. Chen, X. Zhou, W. Zhang, J. Yuan, and J. Feng. 2015. The mechanism underlying dibutyl phthalate induced shortened anogenital distance and hypospadias in rats. J. Pediatr. Surg. 50(12):2078-2083.

Li, Y., M. Zhuang, T. Li, and N. Shi. 2009. Neurobehavioral toxicity study of dibutyl phthalate on rats following in utero and lactational exposure. J. Appl. Toxicol. 29(7):603-611.

Lin, H., R.S. Ge, G.R. Chen, G.X. Hu, L. Dong, Q.Q. Lian, 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. Proc. Natl. Acad. Sci. U.S.A. 105(20):7218-7222.

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. Biol. Reprod. 80(5):882-888.

Lin, L.C., S.L. Wang, Y.C. Chang, P.C. Huang, J.T. Cheng, P.H. Su, and P.C. Liao. 2011. Associations between maternal phthalate exposure and cord sex hormones in human infants. Chemosphere 83(8):1192-1199.

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. J. Expo. Sci. Environ. Epidemiol. 5(4):343-353.

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. J. Toxicol. Environ. Health A 71(19):1289-1294.

Lorber, M., and A.M. Calafat. 2012. Dose reconstruction of di(2-ethylhexyl) phthalate using a simple pharmacokinetic model. Environ. Health Perspect. 120(12):1705-1710.

Lorber, M., J. Angerer, and H.M. Koch. 2010. A simple pharmacokinetic model to characterize exposure of Americans to di-2-ethylhexyl phthalate. J. Expo. Sci. Environ. Epidemiol. 20(1):38-53.

MacLeod, D.J., R.M. Sharpe, M. Welsh, M. Fisken, H.M. Scott, G.R. Hutchison, A.J. Drake, and S. van den Driesche. 2010. Androgen action in the masculinization programming window and development of male reproductive organs. Int. J. Androl. 33(2):279-287.

Mahood, I.K., H.M. Scott, R. Brown, N. Hallmark, M. Walker, and R.M. Sharpe. 2007. In utero exposure to di(n-butyl) phthalate and testicular dysgenesis: Comparison of fetal and adult end points and their dose sensitivity. Environ. Health Perspect. 115(Suppl. 1):55-61.

Martinez-Arguelles, D.B., T. Guichard, M. Culty, B.R. Zirkin, and V. Papadopoulos. 2011. In utero exposure to the antiandrogen di-(2-ethylhexyl) phthalate decreases adrenal aldosterone production in the adult rat. Biol. Reprod. 85(1):51-61.

Martinez-Arguelles, D.B., E. Campioli, M. Culty, B.R. Zirkin, and V. Papadopoulos. 2013. Fetal origin of endocrine dysfunction in the adult: The phthalate model. J. Steroid Biochem. Mol. Biol. 137:5-17.

Martino-Andrade, A.J., R.N. Morais, G.G. Botelho, G. Muller, S.W. Grande, G.B. Carpentieri, G.M. Leao, and P.R. Dalsenter. 2009. Coadministration of active phthalates results in disruption of foetal testicular function in rats. Int. J. Androl. 32(6):704-712.

Martino-Andrade, A.J., F. Liu, S. Sathyanarayana, E.S. Barrett, J.B. Redmon, R.H. Nguyen, H. Levine, and S.H. Swan. 2016. Timing of prenatal phthalate exposure in relation to genital endpoints in male newborns. Andrology 4(4):585-593.

Masutomi, N., M. Shibutani, H. Takagi, C. Uneyama, N. Takahashi, and M. Hirose**.** 2003. Impact of dietary exposure to methoxychlor, genistein, or diisononyl phthalate during the perinatal period on the development of the rat endocrine/reproductive systems in later life. Toxicology 192(2-3):149-170.

McIntyre, B.S., N.J. Barlow, and P.M. Foster. 2001. Androgen-mediated development in male rat offspring exposed to flutamide in utero: Permanence and correlation of early postnatal changes in anogenital distance and nipple retention with malformations in androgen-dependent tissues. Toxicol. Sci. 62(2):236-249.

McKinnell, C., R.T. Mitchell, M. Walker, K. Morris, C.J. Kelnar, W.H. Wallace, and R.M. Sharpe. 2009. Effect of fetal or neonatal exposure to monobutyl phthalate (MBP) on testicular development and function in the marmoset. Hum. Reprod. 24(9):2244-2254.

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. J. Clin. Endocrinol. Metab. 97(3):E341-E348.

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. Environ. Health Perspect. 109(3):229-237.

Mylchreest, E., R.C. Cattley, and P.M. Foster. 1998. Male reproductive tract malformations in rats following gestational and lactational exposure to di(n-butyl) phthalate: An antiandrogenic mechanism? Toxicol. Sci. 43(1):47-60.

Mylchreest, E., M. Sar, R.C. Cattley, and P.M. Foster. 1999. Disruption of androgen-regulated male reproductive development by di(n-butyl) phthalate during late gestation in rats is different from flutamide. Toxicol. Appl. Pharmacol. 156(2):81-95.

Mylchreest, E., D.G. Wallace, R.C. Cattley, and P.M. Foster. 2000. Dose-dependent alterations in androgen-regulated male reproductive development in rats exposed to di(n-butyl) phthalate during late gestation. Toxicol. Sci. 55(1):143-151.

Nagao, T., R. Ohta, H. Marumo, T. Shindo, S. Yoshimura, and H. Ono. 2000. Effect of butyl benzyl phthalate in Sprague-Dawley rats after gavage administration: A two-generation reproductive study. Reprod. Toxicol. 14(6):513-532.

NRC (National Research Council). 2006. Assessing the Human Health Effects of Trichloroethylene. Washington, DC: The National Academies Press.

NRC. 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]. Available: https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed June 17, 2016].

Pocar, P., N. Fiandanese, C. Secchi, A. Berrini, B. Fischer, J.S. Schmidt, K. Schaedlich, and V. Borromeo**.** 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(2):937-948.

Raudenbush, S.W. 2009. Analyzing effect sizes: Random-effects models. Pp. 295-315 in The Handbook of Research Synthesis and Meta-Analysis, 2nd Ed., H. Cooper, L.V. Hedges, and J.C. Valentine, eds. 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. Environ. Health Perspect. 122(7):711-718.

Saillenfait, A.M., J.P. Payan, J.P. Fabry, D. Beydon, I. Langonne, F. Gallissot, and J.P. Sabate. 1998. Assessment of the developmental toxicity, metabolism, and placental transfer of di-n-butyl phthalate administered to pregnant rats. Toxicol. Sci. 45(2):212-224.

Saillenfait, A.M., J.P. Sabate, and F. Gallissot. 2008. Diisobutyl phthalate impairs the androgen-dependent reproductive development of the male rat. Reprod. Toxicol. 26(2):107-115.

Saillenfait, A.M., J.P. Sabate, and F. Gallissot. 2009. Effects of in utero exposure to di-n-hexyl phthalate on the reproductive development of the male rat. Reprod. Toxicol. 28(4):468-476.

Saillenfait, A.M., A.C. Roudot, F. Gallissot, and J.P. Sabate. 2011. Prenatal developmental toxicity studies on di-n-heptyl and di-n-octyl phthalates in Sprague-Dawley rats. Reprod. Toxicol. 32(3):268-276.

Saillenfait, A.M., J.P. Sabate, A. Robert, B. Cossec, A.C. Roudot, F. Denis, and M. Burgart. 2013a. Adverse effects of diisooctyl phthalate on the male rat reproductive development following prenatal exposure. Reprod. Toxicol. 42:192-202.

Saillenfait, A.M., J.P. Sabate, A. Robert, V. Rouiller-Fabre, A.C. Roudot, D. Moison, and F. Denis. 2013b. Dose-dependent alterations in gene expression and testosterone production in fetal rat testis after exposure to di-n-hexyl phthalate. J. Appl. Toxicol. 33(9):1027-1035.

Samandar, E., M.J. Silva, J.A. Reidy, L.L Needham, and A.M. Calafat. 2009. Temporal stability of eight phthalate metabolites and their glucuronide conjugates in human urine. Environ. Res. 109(5):641-646.

Sathyanarayana, S., E. Barrett, S. Butts, C. Wang, and S.H. Swan. 2014. Phthalate exposure and reproductive hormone concentrations in pregnancy. Reproduction 147(4):401-409.

Scarano, W.R., F.C. Toledo, M.T. Guerra, P.F. Pinheiro, R.F. Domeniconi, S.L. Felisbino, S.G. Campos, S.R. Taboga, and W.G. Kempinas. 2010. Functional and morphological reproductive aspects in male rats exposed to di-n-butyl phthalate (DBP) in utero and during lactation. J. Toxicol. Environ. Health A 73(13-14):972-984.

Scott, H.M., J.I. Mason, and R.M. Sharpe. 2009. Steroidogenesis in the fetal testis and its susceptibility to disruption by exogenous compounds. Endocr. Rev. 30(7):883-925.

Sena, E.S., G.L. Currie, S.K. McCann, M.R. Macleod, and D.W. Howells. 2014. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. J. Cereb. Blood Flow Metab. 34(5):737-742.

Silva, M.J., J.A. Reidy, A.R. Herbert, J.L. Preau, L.L. Needham, and A.M. Calafat. 2004. Detection of phthalate metabolites in human aminiotic fluid. Bull. Environ. Contam. Toxicol. 72(6):1226-1231.

Skakkebaek, N.E. 2002. Endocrine disrupters and testicular dysgenesis syndrome. Horm. Res. 57(Suppl. 2):43.

Spade, D.J., E.V. McDonnell, N.E. Heger, J.A. Sanders, C.M. Saffarini, P.A. Gruppuso, M.E. De Paepe, and K. Boekelheid. 2014. Xenotransplantation models to study the effects of toxicants on human fetal tissues. Birth Defects Res. B Dev. Reprod. Toxicol. 101(6):410-422.

Stroheker, T., J.F. Regnier, J. Lassurguere, and M.C. Chagnon. 2006. Effect of in utero exposure to di-(2-ethylhexyl)phthalate: Distribution in the rat fetus and testosterone production by rat fetal testis in culture. Food Chem. Toxicol. 44(12):2064-2069.

Struve, M.F., K.W. Gaido, J.B. Hensley, K.P. Lehmann, S.M. Ross, M.A. Sochaski, G.A. Willson, and D.C. Dorman. 2009. Reproductive toxicity and pharmacokinetics of di-n-butyl phthalate (DBP) following dietary exposure of pregnant rats. Birth Defects Res. B Dev. Reprod. Toxicol. 86(4):345-354.

Sullivan, S., and J.L. Teague. 2005. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ. Health Perspect. 113(8):1056-1061.

Suzuki, Y., J. Yoshinaga, Y. Mizumoto, S. Serizawa, and H. Shiraishi. 2012. Foetal exposure to phthalate esters and anogenital distance in male newborns. Int. J. Androl. 35(3):236-244.

Swan, S.H. 2008. Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. Environ. Res. 108(2):177-184.

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, and J.L. Teague. 2005. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ. Health Perspect. 113(8):1056-1061.

Swan, S.H., S. Sathyanarayana, E.S. Barrett, S. Janssen, F. Liu, R.H. Nguyen, and J.B. Redmon. 2015. First trimester phthalate exposure and anogenital distance in newborns. Hum. Reprod. 30(4):963-972.

Thankamony, A., N. Lek, D. Carroll, M. Williams, D.B. Dunger, C.L. Acerini, K.K. Ong, and I.A. Hughes. 2014. Anogenital distance and penile length in infants with hypospadias or cryptorchidism: Comparison with normative data. Environ. Health Perspect. 122(2):207-211.

Thompson, C.J., S.M. Ross, J. Hensley, K. Liu, S.C. Heinze, S.S. Young, and K.W. Gaido. 2005. Differential steroidogenic gene expression in the fetal adrenal gland versus the testis and rapid and dynamic response of the fetal testis to di(n-butyl) phthalate. Biol. Reprod. 73(5):908-917.

Tyl, R.W., C.B. Myers, M.C. Marr, P.A. Fail, J.C. Seely, D.R. Brine, R.A. Barter, and J.H. Butala. 2004. Reproductive toxicity evaluation of dietary butyl benzyl phthalate (BBP) in rats. Reprod. Toxicol. 18(2):241-264.

van den Driesche, S., P. Kolovos, S. Platts, A.J. Drake, and R.M. Sharpe. 2012. Inter-relationship between testicular dysgenesis and Leydig cell function in the masculinization programming window in the rat. PLOS One 7(1):e30111.

van der Zanden, L.F., I.A. van Rooij, W.F. Feitz, B. Franke, N.V. Knoers, and N. Roeleveld. 2012. Aetiology of hypospadias: A systematic review of genes and environment. Hum. Reprod. Update 18(3):260-283.

Viechtbauer, W. 2005. Bias and efficiency of meta-analytic variance estimators in the random-effects model. J. Educ. Behavior. Stat. 30(3):261-293.

Viechtbauer, W. 2010. Conducting meta-analysis in R with the metaphor package. J. Stat. Softw.36(3):1-48.

Vo, T.T., E.M. Jung, V.H. Dang, K. Jung, J. Baek, K.C. Choi, and E.B. Jeung. 2009. Differential effects of flutamide and di-(2-ethylhexyl) phthalate on male reproductive organs in a rat model. J. Reprod. Dev. 55(4):400-411.

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. J. Clin. Invest. 118(4):1479-1490.

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. Toxicol. Lett. 170(3):177-184.

Wilson, V.S., C.R. Blystone, A.K. Hotchkiss, C.V. Rider, and L.E. Gray, Jr. 2008. Diverse mechanisms of anti-androgen action: Impact on male rat reproductive tract development. Int. J. Androl. 31(2):178-187.

Wittassek, M., and J. Angerer. 2008. Phthalates: Metabolism and exposure. Int. J. Androl. 31(2):131-138.

Wittassek, M., J. Angerer, M. Kolossa-Gehring, S.D. Schafer, W. Klockenbusch, L. Dober, A.K. Gunsel, A. Muller, and G.A. Wiesmuller. 2009. Fetal exposure to phthalates—a pilot study. Int. J. Hyg. Environ. Health 212(5):492-498.

Wohlfahrt-Veje, C., K.M. Main, and N.E. Skakkebaek. 2009. Testicular dysgenesis syndrome: Foetal origin of adult reproductive problems. Clin. Endocrinol. 71(4):459-465.

Wolfe, G.W., and K.A. Layton. 2005. Diethylhexylphthalate: Multigenerational Reproductive Assessment by Continuous Breeding when Administered to Sprague-Dawley Rats in the Diet. TRC Study No. 7244-200. NTP-RACB 98-004. TherImmune Research Corporation, Gaithersburg, MD.

Wolfe, G.W., and R.V. Patel. 2002. Dibutyl Phthalate: Multigenerational Reproductive Assessment by Continuous Breeding When Administered to Sprague-Dawley Rats in the Diet, Vol.1. NTP RACB 97003. TherImmune Research Corporation, Gaithersburg, MD.

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. Biol. Res. 46(2):139-146.

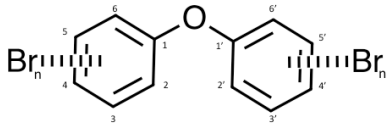
Zhang, Y., X. Jiang, and B. Chen. 2004. Reproductive and developmental toxicity in F1 Sprague-Dawley male rats exposed to di-n-butyl phthalate in utero and during lactation and determination of its NOAEL. Reprod. Toxicol. 18(5):669-676.

Zota, A.R., A.M. Calafat, and T.J. Woodruff. 2014. Temporal trends in phthalate exposure: Findings from the National Health and Nutrition Examination Survey, 2001-2010. Environ. Health Perspect. 122(3):235-241.

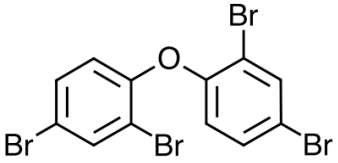
4

Effect of Polybrominated Diphenyl Ethers on Neurodevelopment

Polybrominated diphenyl ethers (PBDEs) are a class of brominated hydrocarbons that have been used as flame retardants in a variety of products, such as textiles, plastics, electronic materials, and polyurethane foams for furniture (EPA 2010). The class is comprised of 209 congeners that share a brominated diphenyl ether molecule with up to 10 bromine atoms attached (see Figure 4-1). Three classes of commercial formulations of PBDE mixtures were produced—the pentaPBDEs, octaPBDEs, and decaPBDEs—but they are no longer produced or used. They are ubiquitous in the environment; they have been shown to be persistent and to bioaccumulate; and human exposure to them has been well documented (Bramwell et al. 2016).



Generic polybrominated diphenyl ether (PBDE)



2,2′,4,4′-Tetrabromodiphenyl ether

FIGURE 4-1 Generic structure of a polybrominated diphenyl ether (PBDE). The position and number of bromine atoms (n) can vary within this class of chemical. A representative PBDE, 2,2′,4,4′-tetrabromodiphenyl ether (BDE-47), is also shown.

PBDEs can undergo a variety of metabolic changes depending on their structure and the degree of bromine substitution (Stapleton et al. 2009). For example, rodents given 2,2′,4,4′,5-penta-bromodiphenyl ether (BDE-99) produce hydroxylated congeners (OH-BDE) and other metabolites (Chen et al. 2006; Qiu et al. 2007). BDE-99 undergoes more extensive metabolism when compared with either BDE-47, 2,2′,4,4′,5,5′-hexabromodiphenyl ether (BDE-153), or 2,2′,3,3′,4,4′,5,5′,6,6′-deca-bromodiphenyl ether (BDE-209) (Chen et al. 2006; Staskal et al. 2006). Glucuronide and sulfate conjugates of dibromophenols and tribromophenols are also formed and excreted in urine and feces (Ho et al. 2015).

PBDEs have been measured in the air, soil, dust, and various foods (EPA 2010). The predominant congeners evaluated and detected are BDE-47, BDE-99, and BDE-153. One approach to monitor human exposure to PBDEs involves measurement of BDE, OH-BDEs, and methoxylated (MeO-BDEs) congeners in blood, milk, and other biological samples (EPA 2010; Wiseman et al. 2011; Butryn et al. 2015). The apparent half-lives of octaBDEs and decaBDEs in human serum are less than 3 months (Thuresson et al. 2006), whereas the half-lives for pentaBDEs may be 2 years or more (Geyer et al. 2004).

PBDEs have been linked to effects on the liver and the nervous system (EFSA 2011; Linares et al. 2015). Concerns related to possible endocrine-disrupting effects associated with PBDEs and OH-BDEs also exist (EFSA 2011; Linares et al. 2015). The OH-BDEs have greater endocrine activity than the parent PBDE does (Ptak et al. 2005, 2006). Certain PBDEs and OH-BDEs alter estrogen receptor alpha and estrogen receptor beta activity in cultured cells (Meerts et al. 2001), rats (Ceccatelli et al. 2006), and mice (Mercado-Feliciano and Bigsby 2008). In vitro, OH-BDEs also affect aromatase (CYP19) and steroid 17α-monooxygenase (CYP17) activities, two enzymes involved in estrogen and androgen synthesis (Canton et al. 2005; Canton et al. 2008; He et al. 2008; Karpeta et al. 2013). Changes in female sexual behavior and estrous cyclicity and alterations in expression patterns of estrogen-regulated genes in sexually dimorphic brain regions occur in rats following in utero exposure to BDE-99 (Lilienthal et al. 2006; Faass et al. 2013). The PBDEs also affect thyroid hormone homeostasis in a number of ways. OH-BDEs share structural similarities with thyroid hormones and bind thyroid hormone receptors and serum thyroid hormone binding proteins and lower free and total thyroxine (T4) concentrations in animals (Fowles et al. 1994; Darnerud and Sinjari 1996; Meerts et al. 2000; Siddiqi et al. 2003; Linares et al. 2015).

PBDE effects on the developing brain are receiving considerable attention in the public-health community. The possible association between PBDE exposure and neurodevelopmental effects was originally considered for a topic at the committee's workshop on February 3, 2016, which was designed to explore potential case studies and to help the committee select the topics for its systematic reviews. During the process of planning the workshop, the committee became aware of a systematic review on the topic that was already under way by Lam et al. (2015, 2016). The focus of the Lam et al. review was the possible association between PBDE exposure and measures of intelligence or attention deficit/hyperactivity disorder (ADHD) and attention-related behavioral conditions in children. The committee judged that a systematic review of animal studies on PBDE exposure and neurodevelopmental outcomes would be an appropriate complement to that review of the human literature. Because the committee's goal was to identify end points that could be considered relevant for human intelligence and ADHD and other attention-related behavioral conditions, the committee's systematic review of the animal data focused on PBDE studies that measured learning, memory, attention, and response inhibition. Animal studies that solely evaluated motor function, fear conditioning, and other tests that were not directly linked to learning, memory, attention, or response inhibition were considered outside the scope of the systematic review. This chapter describes the systematic review of the animal evidence on PBDE neurodevelopmental effects that the committee performed. It also uses the review of human evidence to demonstrate how an existing systematic review can be critically evaluated, used, and updated. The separate lines of evidence are subsequently integrated with mechanistic information to draw conclusions about associations.

SYSTEMATIC REVIEW METHODS

The systematic reviews were designed to answer two related questions:

- Is developmental exposure to PBDEs in nonhuman mammals associated with alterations in learning, memory, attention, or response inhibition? This question was addressed through the conduct of an independent systematic review of the relevant animal literature.
- Is developmental exposure to PBDEs in humans associated with alterations in quantitative measures of intelligence or ADHD and attention-related behavioral conditions? The committee updated a recent systematic review of the human literature by Lam et al. (2015, 2016) to address the question.

The PECO (Population, Exposure, Comparator, and Outcome) statements for the systematic reviews of the animal and the human studies are presented in Boxes 4-1 and 4-2, and the protocols used to conduct the systematic reviews are provided in Appendix E (Section E-1) and Appendix F (Section F-1), respectively. The protocols were peer reviewed in accordance with the standard report-review practices of the National Academies of Sciences, Engineering, and Medicine. Most of the peer reviewers of the protocols

were also peer reviewers of this report to ensure that the original protocols were followed and that any revisions or updates have been appropriately documented and justified. (See the Acknowledgments for the list of peer reviewers.) A summary of the methods is briefly described below.

BOX 4-1 PECO Statement for the PBDE (Animal) Systematic Review

Population: Nonhuman mammals

Exposure:

- PBDE refers to any single PBDE congener or combination of grouped congeners.
- Any developmental exposure to PBDEs, with no restrictions based on route of exposure or administered dose or concentration. To be considered "developmental," the exposure occurred during any of the following periods: prior to conception in one or both parents, prenatal in the pregnant female (exposure to offspring in utero), or postnatal until sexual maturation.

Comparator: Nonhuman mammals exposed during development to different doses of PBDEs or vehicle-only treatment.

Outcomes: Measures of learning, memory, attention, or response inhibition. Examples of tests include Morris water maze, radial arm maze, and operant tests of cognition.

BOX 4-2 PECO Statement for the PBDE (Human) Systematic Review

Population: Humans without restriction based on age

Exposure:

- PBDE refers to any single PBDE congener or combination of grouped congeners.
- Developmental exposure to PBDEs. To be considered developmental, the exposure occurred during any of the following: prior to conception for one or both parents, during pregnancy (exposure to offspring in utero), perinatally, or in childhood.
- Exposure measurements must be from human biological samples (e.g., urine, blood, or other specimens).

Comparator: Humans exposed to lower levels of PBDEs.

Outcomes:

- Quantitative measures of intelligence. For example, measures from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Wechsler Intelligence Scale for Children (WISC), Stanford-Binet Intelligence Scale, or the McCarthy Scales of Children's Abilities (MSCA).
- Outcome measures of ADHD and attention-related behavioral conditions. For example, measures from the Child Behavior Checklist (CBCL)/1.5-5, Conners' Kiddie Continuous Performance Test (K-CPT), Conners' Rating Scale-Teachers (CRS-T), Conners' Parent Rating Scale-Revised (CPRS), WISC-III (selected subscales), the Disruptive Behavior Disorders Rating Scale (DBD), or Continuous ADHD Confidence Index score.

Literature Searches and Screening

Searches for relevant existing systematic reviews were performed first. PubMed was searched for systematic reviews published in 2013 or later, and the systematic-review protocol registries PROSPERO and CAMARADES were searched for relevant protocols on August 3, 2016. Citations found in searching for systematic reviews of PBDEs and measures of intelligence or ADHD and attention-related behavioral conditions were considered a systematic review if they met the following minimum criteria: (1) conducted an explicit and adequate literature search; (2) applied predefined eligibility criteria; (3) considered the quality of included studies or risk of bias assessment; and (4) synthesized (or attempted to synthesize) the findings, either qualitatively or quantitatively. This definition of a systematic review is consistent with previous Institute of Medicine recommendations that defined a systematic review as "a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

Animal Evidence

For the independent systematic review of the animal evidence, scientific literature databases were searched for relevant studies on the effects of developmental exposure to PBDEs on measures of learning, memory, attention, or response inhibition in nonhuman mammals. A medical librarian, with specific training and expertise in performing searches for systematic reviews, developed and conducted the searches. Searches to support the systematic review of the animal studies were performed by the librarian in PubMed, Embase, and Toxline on August 15, 2016. The search strategies for the animal data are presented in the protocol provided in Appendix E (Section E-1b).

References were screened at the title and abstract level and at the full-text level by the same two reviewers using DistillerSR (https://www.evidencepartners.com). The screening criteria used are specified in the protocol for the animal systematic review (see Appendix E, Section E-1c). At the title and abstract screening level, if there was disagreement between the reviewers or an abstract was not available, the reference was passed on to the full-text screening level for further review. At the full-text level, disagreements about whether to include a study were discussed by the two reviewers to reach consensus; if consensus could not be reached, a third team member was consulted to resolve the differences.

Human Evidence

Literature searches were performed to update the systematic review performed by Lam et al. (2015, 2016). A librarian used the same literature search methods and databases as the existing review to search for new reports on September 28, 2016, and searched from a year before the last search date of the review so that there was a 1-year overlap between the two searches.

References were screened at the title and abstract level and at the full-text level by the same two reviewers using DistillerSR (https://www.evidencepartners.com). The screening criteria used are specified in protocol for the existing human systematic review. At the title and abstract screening level, if there was disagreement between the reviewers or an abstract was not available, the reference was passed on to the full-text screening level for further review. At the full-text level, disagreements about whether to include a study were discussed by the two reviewers to reach consensus; if consensus could not be reached, a third team member was consulted to resolve the differences.

Evaluating a Systematic Review

The Lam et al. (2015, 2016) systematic review was evaluated for risk of bias using the ROBIS tool (Whiting et al. 2016). The tool has three evaluation phases: phase 1 involves assessing the relevance; phase 2 involves identifying concerns with the review process; and phase 3 involves judging risk of bias. In the first phase, the relevance of the systematic review was assessed by comparing the committee's tar-

get question with the question being addressed in the review. The PECO framework was used to assess the match between the target question and the question addressed by the review. This phase was informally completed by the entire committee early in the process (immediately following its workshop in February 2016), and the committee's systematic review protocol was designed to be an update of an existing systematic review.

The next two phases of applying the ROBIS tool involved two committee members who used the tool to identify concerns with the review process and to judge risk of bias in the review. Four domains were used to identify concerns with the review process: (1) study eligibility criteria, (2) identification and selection of studies, (3) data collection and study appraisal, and (4) synthesis and findings. Included for each domain were questions to help assess specific issues with potential biases. The response to the questions helped in judging overall risk of bias for each domain. In the third phase, the overall risk of bias in the review (low, high, or unclear) was determined.

In contrast with the animal systematic review, which followed methods developed by the National Toxicology Program's (NTP's) Office of Health Assessment and Translation (OHAT), the Lam et al. (2015, 2016) review of the human evidence used the Navigation Guide methodology (Woodruff and Sutton 2014). Figure 4-2 illustrates the steps of the Navigation Guide method. The two approaches are very similar (see Table 1-1 in Chapter 1 for a side-by-side comparison), and they are based on the same established methodology for the conduct of systematic review and evidence assessment (e.g., Cochrane Collaboration, AHRQ Evidence-based Practice Center Program, and GRADE). Both the OHAT and Navigation Guide methods include the key steps recommended by a previous National Academies committee (NRC 2014) for problem formulation, protocol development, specifying a study question, developing a PECO statement, identifying and selecting the evidence, evaluating the evidence, and integrating the evidence. Different terminology, however, is used to describe the results of analyzing and integrating the evidence. Consequently, it was necessary for the committee to "translate" the findings from the Lam et al. (2015, 2016) systematic review into OHAT ratings for the evidence integration (discussed later in this section).

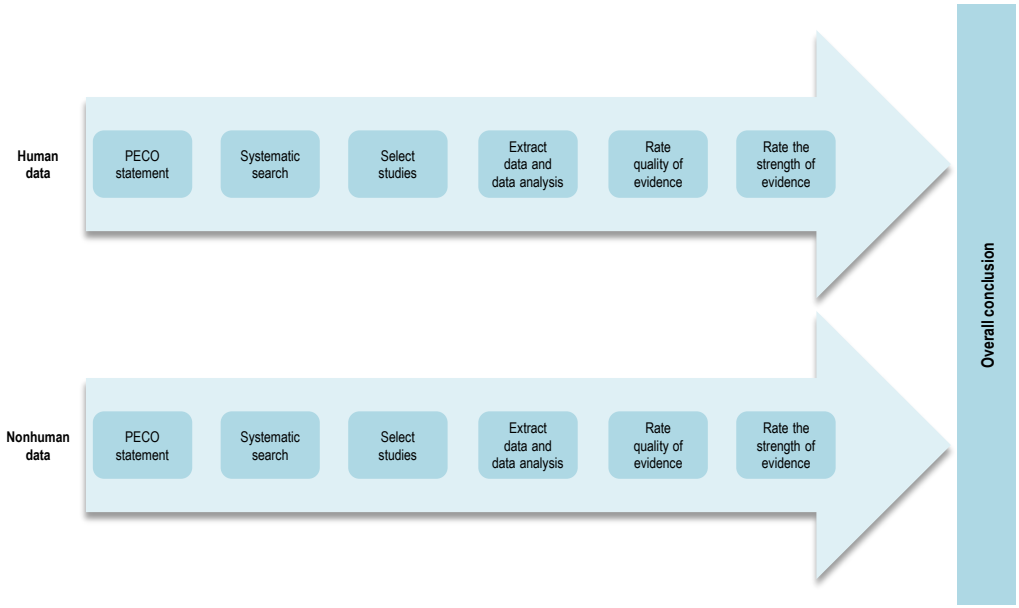


FIGURE 4-2 Steps in the Navigation Guide protocol. SOURCE: Woodruff and Sutton (2014).

Data Extraction

Animal data from the included studies were entered into the Health Assessment Workspace Collaborative (HAWC), a Web-based interface application for warehousing data and creating visualizations (https://hawcproject.org). See Appendix E, Section E-1d, for data extraction elements for the animal studies. One person entered data and a second person verified the entries. All data entered into HAWC are available at the following link: https://hawcproject.org/assessment/352/. For the human evidence, new information was summarized and considered in context with the results of the existing systematic review.

Risk of Bias and Study Quality Evaluations

Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect. Internal validity or risk of bias was assessed for individual animal studies using the OHAT method that outlines an approach to evaluating risk of bias for experimental animal studies (NTP 2015). The criteria were customized from the basic OHAT method and described in the protocol for addressing the specific research question for this review (e.g., appropriate methods for PBDE exposure characterization and use of litter as the unit of analysis) (Appendix E, Section E-1e). Key risk of bias elements for animal studies included reliability of the outcome measure, blinding of researchers to treatment groups, and whether investigators controlled for litter effects in their experimental design or statistical approaches. Two committee members independently assessed each study and answered all applicable risk of bias questions following prespecified criteria detailed in the study protocol. One individual of the pair then reconciled any discrepancies with input from the second committee member.

For the human studies, the committee accepted the risk of bias approach used in the Lam et al. (2015, 2016) review. The review was performed similarly and included risk of bias domains very similar to those used by NTP (2015). The Navigation Guide examines nine risk of bias domains: source population representation, blinding, exposure assessment, outcome assessment, confounding, incomplete outcome data, selective outcome reporting, conflict of interest, and other sources of bias (Woodruff and Sutton 2014). Similar to the OHAT method, each domain was rated as low, probably low, probably high, high, or not applicable.

Data Analysis and Evidence Integration

For the evaluation of the animal evidence, the body of evidence on each outcome was synthesized qualitatively and, where appropriate, a meta-analysis was performed. Summaries of main characteristics for each included study were compiled and reviewed by two committee members to determine comparability between studies, to identify data transformations necessary to ensure comparability, and to determine whether heterogeneity was a concern. The main characteristics considered across all eligible studies included the following:

- Experimental design (e.g., acute, chronic, multigenerational);
- Animal model used (e.g., species, strain, sex, genetic background);
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status);
- Developmental stage of animals at treatment and outcome assessment;
- Dose levels, frequency of treatment, timing, duration, and exposure route;
- Health outcome(s) and specific measures reported;
- Type of data (e.g., continuous or dichotomous), statistics presented in paper; and
- Variation in the degree of risk of bias at individual study level.

For the human evidence, the expanded body of evidence was synthesized qualitatively, and a determination was made about whether the new information would substantially affect the conclusions drawn

in the Lam et al. (2015, 2016) systematic review. If the data were determined to materially affect the evidence base, any quantitative evaluations performed in the original review were updated.

Confidence Rating and Level of Evidence Conclusions

For the animal systematic review, the confidence in the body of evidence for each outcome was evaluated using a grading system based on a modification of the GRADE system for rating certainty in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). The process for rating confidence in the body of evidence as high, moderate, low, or very low was guided by the *OHAT Handbook for Conducting a Literature-Based Health Assessment* (NTP 2015) (see Figure 4-3). In brief, studies on a particular outcome were initially grouped by key study design features, and each grouping of studies was given an initial confidence rating by those features. Several factors were then considered to determine whether the initial rating should be downgraded or upgraded. Factors that decrease confidence in results and lead to downgrading are risk of bias, unexplained inconsistency in results, indirectness or lack of applicability, imprecision, and publication bias. Factors that increase confidence in results and can upgrade a rating are a large magnitude of effect; evidence of a dose-response relationship; consistency across study designs, populations, animal models, or species; consideration of residual confounding; and other factors that increase confidence in the association or effect (e.g., rare outcomes). Confidence ratings were independently assessed by two committee members, and discrepancies were resolved by consensus and consultation with a third committee member as needed. After a final confidence rating is determined, the rating is translated into a level of evidence using the scheme presented in Figure 4-4.

The Navigation Guide methodology was used by Lam et al. (2015, 2016) to evaluate the body of evidence from human studies. The approach is similar to the OHAT method in that both methods use the GRADE approach to evaluating confidence in the body of evidence; many groups use different terminology for this step, however, and the terminology has changed from "quality" to "confidence" and most recently to "certainty" within the GRADE framework (Morgan et al. 2016; Rooney et al. 2017). The Navigation Guide uses the term "quality" of the body of evidence, whereas OHAT method uses the term "confidence" in the body of evidence to reflect the GRADE-based evaluation. The basis for determining the ratings is similar in the two methods and involves a similar process of giving an initial rating to the evidence and then considering factors that could upgrade or downgrade the rating. All human observational studies start out with an initial moderate rating in the Navigation Guide method, and the initial confidence rating assigned to studies in the OHAT method depends on aspects of study design (see Figure 4-3). Cohort and case-control studies would also start out as moderate under the OHAT method, but cross-sectional studies would start at low initial confidence because the study design cannot assure that exposure preceded outcomes. Both methods use the same eight factors from GRADE to consider potential upgrades or downgrades to the body of evidence (risk of bias, unexplained inconsistency, indirectness, imprecision, publication bias, large magnitude of effect, dose response, and residual confounding). The OHAT method also considers two additional potential upgrades for consistency across study designs or diverse populations and for other factors, such as particularly rare outcomes. Although these factors are not part of the Navigation Guide method for rating quality/certainty, they are considered in the next step, so the methods differ in the sequence of consideration of these factors rather than any difference in approach.

After rating the quality/confidence in the body of evidence, a determination is made about the "strength" of the evidence using the Navigation Guide (see Figure 4-2) and about the "level" of evidence using the OHAT method (see Figure 4-4). Both determination schemes consider the GRADE quality/confidence rating in the body of evidence. Then the Navigation Guide also evaluates the likelihood that a new study would change the conclusion and other compelling attributes of the data that might influence certainty. Because the systematic review conducted by the committee on the animal evidence followed the OHAT method, the Lam et al. (2015, 2016) review is described below using the closest equivalent OHAT terminology.

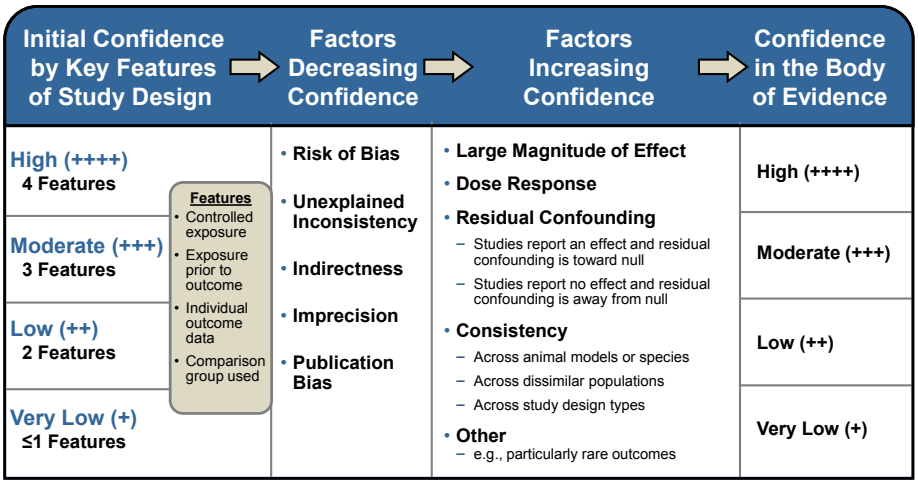
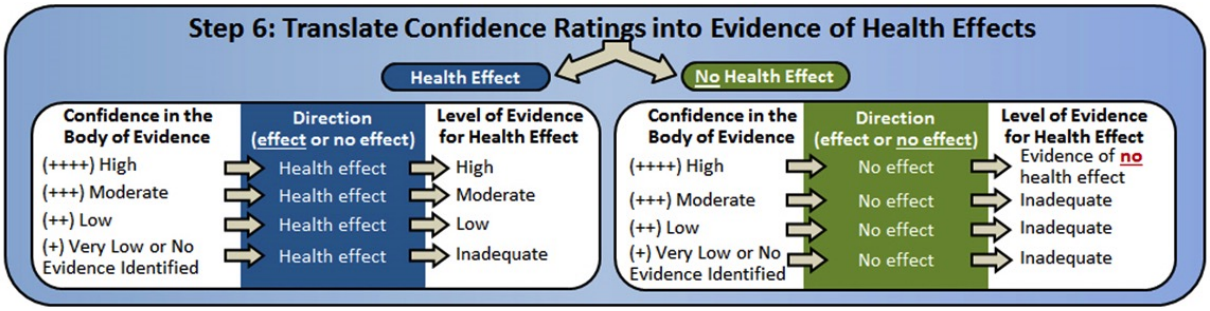


FIGURE 4-3 Method for assessing confidence in the body of evidence. SOURCE: NTP (2015).



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between exposure to the substance and the health outcome(s).
Moderate 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).
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.
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).
Inadequate Evidence	There is insufficient evidence available to assess if exposure to the substance is associated with the health outcome(s).

FIGURE 4-4 Method for translating confidence ratings into evidence for health effects. SOURCE: NTP (2015).

Integration of Evidence and Drawing Hazard Identification Conclusions

The committee used guidance from OHAT to draw hazard identification conclusions (NTP 2015). The procedure involves integrating the levels of evidence ratings for the human and the animal data and considering them within the context of mechanistic information. The five possible hazard conclusions are (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 has been described as having inadequate evidence, conclusions are drawn on the basis of a single evidence base. The hazard identification scheme is presented in Figure 4-5.

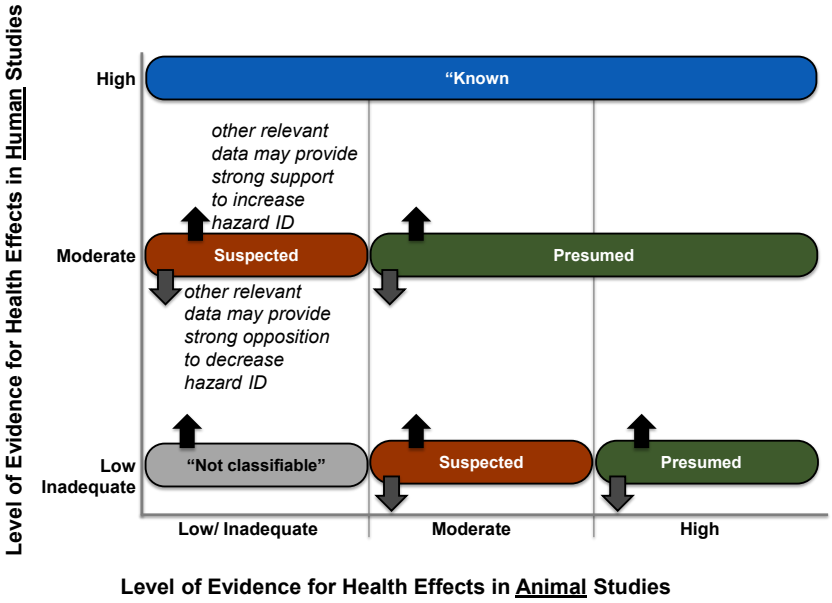


FIGURE 4-5 OHAT hazard identification scheme. SOURCE: NTP (2015).

RESULTS

Literature Search and Screening Results

Animal Studies

A search for recently published systematic reviews on developmental exposure to PBDEs and alterations in learning, memory, attention, or response inhibition in nonhuman mammals found no relevant reviews. No relevant protocols of ongoing reviews were found in PROSPERO or CAMARADES either, so an independent systematic review of the animal literature was performed. A search of PubMed, Embase, and Toxline for relevant publications to address the PECO statement found 1,851 unique citations (see Appendix E, Section E-2, for breakdown by database). A total of 67 publications met the criteria for full-text review, and 27 of those met the inclusion criteria for data extraction (see Figure 4-6). A review of the reference lists of the 27 included studies found no additional publications that were potentially relevant. Thus, 27 publications were included in the review (see Box 4-3).

Human Studies

A search for existing systematic reviews on developmental exposure to PBDEs and effects on intelligence or attention-related conditions in humans found 18 reports (see Appendix F, Section F-2). After screening at the title and abstract level, two reports were evaluated at the full-text level (Roth and Wilks 2014; Lam et al. 2015). One publication was found in the search of PubMed (Roth and Wilks 2014), but it did not meet the criteria established for an appropriate and relevant systematic review because the literature search was restricted to articles published since January 1, 2006, and no formal risk of bias assessment of the studies was performed. A relevant systematic review protocol was found in PROSPERO (Lam et al. 2015); this was the same review that the committee identified in preparing for its workshop. The authors of this systematic review provided the committee with an interim draft of their systematic review (Lam et al. 2016).

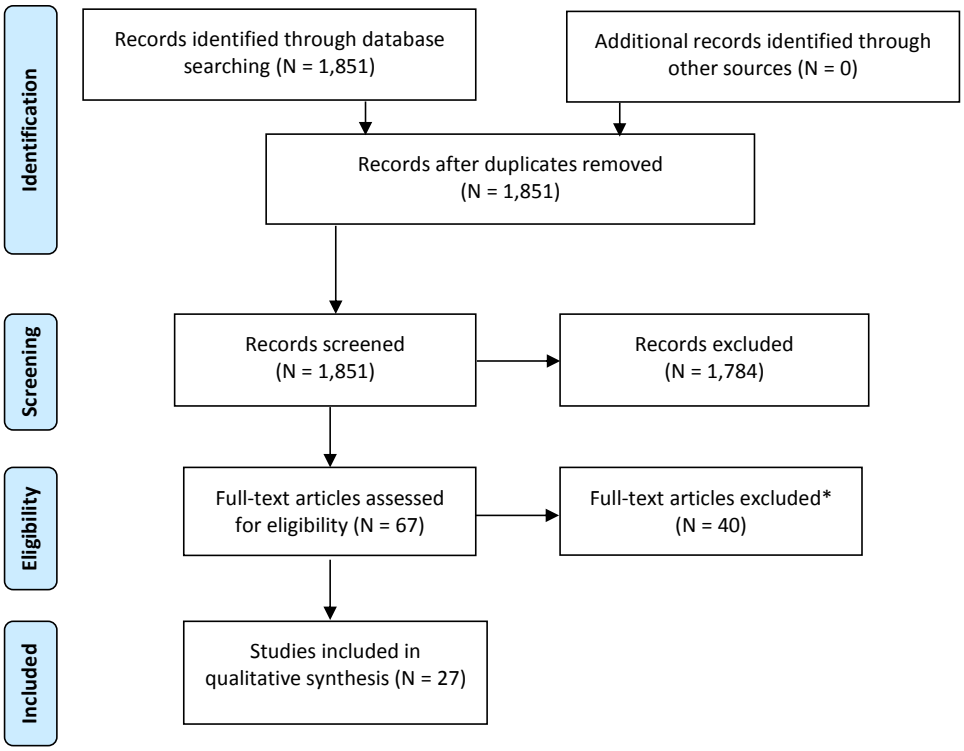


FIGURE 4-6 Summary of the search and screening of the literature on the effects of developmental exposure to PBDEs on learning, memory, attention, or response inhibition in animals. *Articles were excluded for the following reasons: no original data (n = 2); study did not report experimental PBDE exposure (n = 1); study did not include developmental exposure (n = 1); study did not assess or report quantitative measures of learning, memory, attention, or response inhibition (n = 9); not in English (n = 1); or other reason (n = 38). Explanations cited for exclusion because of other reasons included study evaluated behavioral end points, study evaluated motor activity, duplicate, abstract, and erratum. NOTE: the number of studies does not equal the total in the figure because the screeners sometimes excluded a study for different reasons.

BOX 4-3 Studies Included in the PBDE (Animal) Systematic Review

Biesemeier et al. 2011	Dufault et al. 2005	Rice et al. 2009
Blanco et al. 2013	Eriksson et al. 2001	Ta et al. 2011
Bowers et al. 2015	Fischer et al. 2008	Verma et al. 2013
Buratovic et al. 2014	He et al. 2009	Verma et al. 2014
Chen et al. 2014	He et al. 2011	Viberg et al. 2003
Cheng et al. 2009	Koenig et al. 2012	Viberg et al. 2006
de-Miranda et al. 2016	Llansola et al. 2009	Woods et al. 2012
Driscoll et al. 2009	Reverte et al. 2013	Zhang et al. 2013
Driscoll et al. 2012	Reverte et al. 2014	Zhao et al. 2014

For the update of the Lam et al. (2015, 2016) review, a search of BIOSIS, Embase, PubMed, Tox-Net/DART, and Web of Science and relevant websites found 19 publications that met the criteria for full-text review, and three of those met the inclusion criteria (see Figure 4-7). The three reports (Cowell et al. 2015; Sagiv et al. 2015; Zhang et al. 2017) were follow-up assessments of three cohorts already included in the Lam et al. (2015, 2016) review.

Evaluation of the Lam et al. (2015, 2016) Review

The ROBIS tool (Whiting et al. 2016) was used to assess the protocol of the Lam et al. (2015, 2016) review. Two committee members independently assessed the risk of bias for the review and were in agreement in their evaluations. Consequently, it was unnecessary to consult a third committee member to achieve consensus. For each of the four domains, and for the overall assessment of the review, the committee members agreed that the risk of bias was low. No concerns were raised for any of the domains. For instance, there was a comprehensive search, with predefined eligibility criteria applied by two independent committee members.

Two minor issues were identified by the committee members, but neither was judged sufficient to change the overall evaluation. One issue was the possible need to consider updating the risk of bias assessment in the systematic review to the ROBINS-I tool (Sterne et al. 2016). This tool is new, however, and was released only recently, so the committee judged that it should not affect the "data collection and study appraisal". The second issue was in the choice of model used in the meta-analyses. The systematic review used the DerSimonian-Laird estimator, and concerns have been raised recently about the model potentially providing biased estimates (Cornell et al. 2014). Currently, however, there is no agreement as to the most appropriate model to use for random effects meta-analysis, so the use of this model did not impact the assessment for the "synthesis and findings" domain.

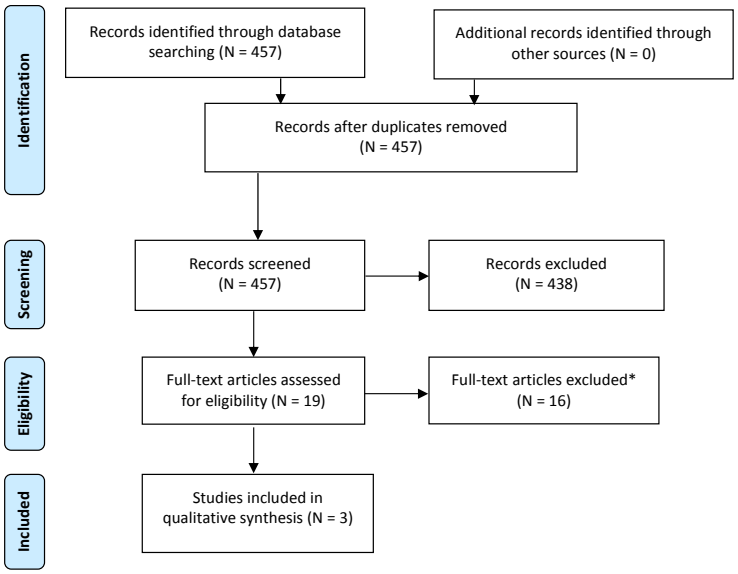


FIGURE 4-7 Summary of the updated search and screening of the literature on the effects of developmental exposure to PBDEs on intelligence or ADHD and attention-related behavioral conditions in humans. *Articles were excluded for the following reasons: study did not include human subjects (n = 1); study did not quantify developmental exposure to PBDE as concentration measured in human biological samples (n = 1); study did not report quantitative measure of intelligence, ADHD, or attention-related behavioral problems (n = 4); duplicate study (n = 4); or other reason (n = 7). Explanations cited for exclusion because of other reasons included study does not evaluate the PBDEs of interest, study is cited in the existing review, and abstract. NOTE: The number of studies does not equal the total in the figure because the screeners sometimes excluded a study for different reasons.

Health Effects Results

Animal Health Effects Results

A wide range of behavioral tests was used to evaluate learning, memory, and attention in rodents exposed to PBDEs during development (see Table 4-1). The most commonly used test of learning and memory in association with PBDEs was the Morris water maze. In this test, animals are typically placed in a large circular pool of water and required to escape by finding and climbing onto a hidden platform. A variety of behaviors can be assessed using this task, including spatial memory (often assessed using a probe trial in which the platform is removed from the pool) or reversal learning (the platform is moved to a different quadrant of the pool). Outcome measures can include escape latency across trials, path length, swim speed, and animal orientation in relation to the platform location. For the purposes of this review, acquisition and reversal learning were considered tests of learning whereas performance in the probe trial was considered an assessment of memory.

Other behavioral tests of learning and memory included the Barnes spatial maze (Koenig et al. 2012); passive avoidance (Zhang et al. 2013); Y maze (Llansola et al. 2009); water T maze (Biesemeier et al. 2011); radial arm maze (Fischer et al. 2008; Verma et al. 2013; de-Miranda et al. 2016); visual discrimination (Dufault et al. 2005; Rice et al. 2009); and operant conditioning test paradigms (Rice et al. 2009). Some studies investigated attention (Driscoll et al. 2009, 2012). For example, Driscoll et al. (2009) provided visual cues with a variable (0-6 sec) pre-cue delay and, in some experiments, also a variable cue duration of 200 to 800 ms. These tasks required the animal to sustain visual attention across five nose poke portals for an indeterminate period of time. There was considerable variability in the animal models used: rats of the Sprague-Dawley, Wistar, and Long Evans strains; mice of the NMRI, C57Bl/6J, apoE2, apoE3, apoE4, Swiss albino, and Mecp2 308+/– strains. The duration of exposure was also highly variable and included acute (single-day) and repeated exposures during gestation and lactation. In general, standardized test methods were not used across studies.

Studies relevant to the systematic review were available on six BDE congeners (BDE-47, -99, -153, -203, -206, and -209) and one technical grade flame retardant mixture (DE-71; a mixture of 24 BDE congeners [Konstantinova et al. 2008]). Studies on learning were available for all these BDEs; studies of memory were available on six of them (BDE-47, -99, -153, -203, and -209 and DE-71); and a study of attention was available on the one mixture (DE-71). No studies on response inhibition were found for any of the BDEs. A variety of different behavioral tests and test parameters were measured; measurements were taken multiple times a day or over several days; and data were presented in different ways. Because of this heterogeneity, the committee found that the PBDE animal data did not lend themselves to creating useful visualizations in HAWC. Thus, summary tables of the evidence were created from the data sets entered into HAWC for evaluation. Confidence in the bodies of evidence on these end points were evaluated for each of the congeners by considering the number of studies available and evaluating factors that would decrease or increase confidence in evidence. All animal studies started with an initial confidence rating of high, because the exposures were controlled, doses were administered before the outcomes were evaluated, individual outcome data were reported, and a comparison group was used. Factors that would upgrade or downgrade confidence in the body of evidence were then considered to determine a final confidence rating. Documentation of how the evidence was evaluated and how the confidence ratings were determined is provided for BDE-47 in this chapter; for the other BDEs, short summaries are provided and are supported by details in Appendix E, Section E-4.

BDE-47 and Learning and Memory

There is moderate confidence in the body of evidence on developmental exposure to BDE-47 and effects on learning in rodents. Six studies of learning were available (see Table 4-2). Two studies in rats found several indications of decreased learning in the Morris water maze (e.g., prolonged latency to find

TABLE 4-1 Studies Included in the PBDE (Animal) Systematic Review

Study	Chemical	Species (strain)	Life Stage Exposed	Life Stage Assessed	Test(s)	Doses (mg/kg-day)
Eriksson et al. 2001	BDE-47	Mouse (NMRI)	PND 10	5 months	Morris water maze	0, 10.5
He et al. 2009	BDE-47	Rat (Sprague-Dawley)	PND 10	2 months	Morris water maze	0, 1, 5, 10
He et al. 2011	BDE-47	Rat (Sprague-Dawley)	PND 10	2 months	Morris water maze	0, 1, 5, 10
Koenig et al. 2012	BDE-47	Mouse (C57Bl/6J)	4 weeks before breeding to PND 21	8 weeks	Barnes spatial maze	0, 0.03, 0.1, 1
Ta et al. 2011	BDE-47	Mouse (C57BL/6J)	GD 0 - PND 21	8 weeks	Morris water maze	0, 0.03, 0.1, 1
Woods et al. 2012	BDE-47	Mouse (Mecp2 308+/-) Mouse (wild type)	GD 0 - PND 21	PND 50-54	Morris water maze	0, 0.03
Blanco et al. 2013	BDE-99	Rat (Sprague-Dawley)	GD 6 - PND 21	PND 26-35	Morris water maze	0, 1, 2
Cheng et al. 2009	BDE-99	Rat (Sprague-Dawley)	GD 6 - PND 21	PND 34-36	Morris water maze	0, 2
Eriksson et al. 2001	BDE-99	Mouse (NMRI)	PND 10	5 months	Morris water maze	0, 12
Fischer et al. 2008	BDE-99	Mouse (NMRI)	PND 10	4 and 6 months	Morris water maze; radial arm maze	0, 0.8
Llansola et al. 2009	BDE-99	Rat (Wistar)	GD 2-9 or GD 11-19	PND 68–70	Y maze	0, 30 (IP)
Zhao et al. 2014	BDE-99	Rat (Sprague-Dawley)	GD 1 - PND 21	PND 34-36	Morris water maze	0, 0.2
Viberg et al. 2003	BDE-153	Mouse (NMRI)	PND 10	PND 180	Morris water maze	0, 0.45, 0.9, 9
Zhang et al. 2013	BDE-153	Rat (Sprague-Dawley)	PND 10	PND 40 and 70	Morris water maze; passive avoidance	0, 1, 5, 10 (IP)
Viberg et al. 2006	BDE-203	Mouse (NMRI)	PND 3 or 10	PND 90	Morris water maze	0, 16.8
Viberg et al. 2006	BDE-206	Mouse (NMRI)	PND 10	PND 90	Morris water maze	0, 18.5
Biesemeier et al. 2011	BDE-209	Rat (Sprague-Dawley)	GD 6 - PND 21	PND 22, PND 62	Water T maze	0, 1, 10, 100, 1000
Buratovic et al. 2014	BDE-209	Mouse (NMRI)	PND 3	5 and 7 months	Morris water maze	0, 3.4, 7.9
Chen et al. 2014	BDE-209	Rat (Sprague-Dawley)	GD 1-14	PND 25	Morris water maze	0, 10.0, 30, 50
Reverte et al. 2013	BDE-209	Mouse (apoE2) Mouse (apoE3) Mouse (apoE4)	PND 10	PND 120 and 360	Morris water maze	0, 10, 30
Reverte et al. 2014	BDE-209	Mouse (apoE2) Mouse (apoE3) Mouse (apoE4)	PND 10	PND 150-180	Fear conditioning	0, 10, 30
Rice et al. 2009	BDE-209	Mouse (C57BL6/J)	PND 2-15	PND 87 or PND 497	Operant (fixed ratio; fixed interval; visual discrimination)	0, 6, 20

(Continued)

TABLE 4-1 Continued

Study	Chemical	Species (strain)	Life Stage Exposed	Life Stage Assessed	Test(s)	Doses (mg/kg-day)
Verma et al. 2013	BDE-209	Mouse (Swiss albino)	PND 3-10	PND 60-66	Morris water maze; radial arm maze	0, 20
Verma et al. 2014	BDE-209	Mouse (Swiss albino)	PND 3-10	NR	Morris water maze	0, 20
Bowers et al. 2015	DE-71	Rat (SpragueDawley)	GD 1 - PND 21	PND 235	Morris water maze	0, 0.3, 3, 30
de-Miranda et al. 2016	DE-71	Rat (Wistar)	PND 5-22	PND 100	Radial maze learning	0, 30
Driscoll et al. 2009	DE-71	Rat (Long-Evans)	GD 0 - PND 21	PND 40-95	Visual discrimination; attention task	0, 3 or 0, 4.5
Driscoll et al. 2012	DE-71	Rat (Long-Evans)	PND 6-12	PND 40-95	Visual task; attention task	0, 5, 15
Dufault et al. 2005	DE-71	Rat (Long-Evans)	PND 6-12	PND 30	Visual discrimination; attention task	0, 30

NOTES: Unless otherwise noted the studies involved oral exposure. GD, gestation day; IP, intraperitoneal; PND, postnatal day; NR, not reported.

the platform) after developmental exposure to BDE-47. Both studies were from the same laboratory (He et al. 2009, 2011). Three of the four mouse studies reported decreased learning in at least one test, strain, or sex and were conducted by different research groups (Eriksson et al. 2001; Ta et al. 2011; Koenig et al. 2012; Woods et al. 2012). The mouse results were variable depending on the tests administered, and a clear pattern was not identified to explain the heterogeneity in response relative to a susceptible strain, to sex, or to dose.

Confidence in the evidence was determined by considering factors that might upgrade or downgrade confidence (see Figure 4-3 for the factors). Confidence was downgraded because of risk-of-bias concerns; all studies had ratings of probably high risk or definitely high risk of bias in at least one of the key issues considered (e.g., lack of randomization of treatment), and most of the studies had multiple risk of bias issues, including not controlling for litter effects in the study design or analysis (see Figure 4-8). Qualitatively, the studies of learning in rodents appeared to have inconsistent results that might warrant a downgrade in confidence because of unexplained inconsistency. Nevertheless, a meta-analysis (presented later in this chapter) of rodent studies on several BDEs, including BDE-47, and latency in the last learning trial of the Morris water maze showed consistent evidence of an association between developmental exposure to PBDEs and decrements in this one measure. Results of the meta-analysis provided another line of evidence to support the consistency of the evidence. Specifically, the meta-analysis found that heterogeneity among studies was low; therefore, the apparent qualitative inconsistency can be explained by differences in precision across studies. No downgrades or upgrades on other factors were made.

There is low confidence in the body of evidence on developmental exposure to BDE-47 and effects on memory in rodents. There were five studies of BDE-47 that tested memory in rodents (see Table 4-3). The one study in rats (He et al. 2011) reported decreased memory in the Morris water maze. Three of the mouse studies (Eriksson et al. 2001; Ta et al. 2011; Koenig et al. 2012) reported no effects on memory, and a fourth study by Woods et al. (2012) reported decrements in memory in female Mecp2 308+/- mice, with no effects on males or in C57BL6 mice of either sex. The data set is similar and contains some of the same studies discussed above with respect to effects of BDE-47 on learning; however, there were fewer studies that reported an effect and one less study overall. Confidence in the body of evidence on memory was downgraded because of risk of bias concerns. All the studies had a probably high risk of bias rating for at least one major issue (e.g., researchers were not blinded to the study groups during outcome assessment), and most of the studies had multiple risk of bias issues, including not controlling for litter effects in the study design or analysis (see Figure 4-9). Confidence was further downgraded for unexplained inconsistencies in the evidence on memory. No downgrades or upgrades on other factors were made.

TABLE 4-2 Studies of BDE-47 and Learning in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
He et al. 2009	Sprague- Dawley rats	PND 10	2 months	Morris water maze	None	1
He et al. 2011	Sprague- Dawley rats	PND 10	2 months	Morris water maze	None	1
Koenig et al. 2012	C57BL/6J mice	GD 0 - PND 21	2 months	Barnes maze	None	0.03
Ta et al. 2011	C57BL/6J mice	GD 0 - PND 21	2 months	Morris water maze	0.1	1
Woods et al. 2012	Female Mecp2 308+/- mice	GD 0 - PND 21	PND 50-53	Morris water maze	None	0.03
	Male Mecp2 308+/- mice	GD 0 - PND 21	PND 50-53	Morris water maze	0.03	None
	Female C57Bl6 mice	GD 0 - PND 21	PND 50-53	Morris water maze	0.03	None
	Male C57Bl6 mice	GD 0 - PND 21	PND 50-53	Morris water maze	0.03	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

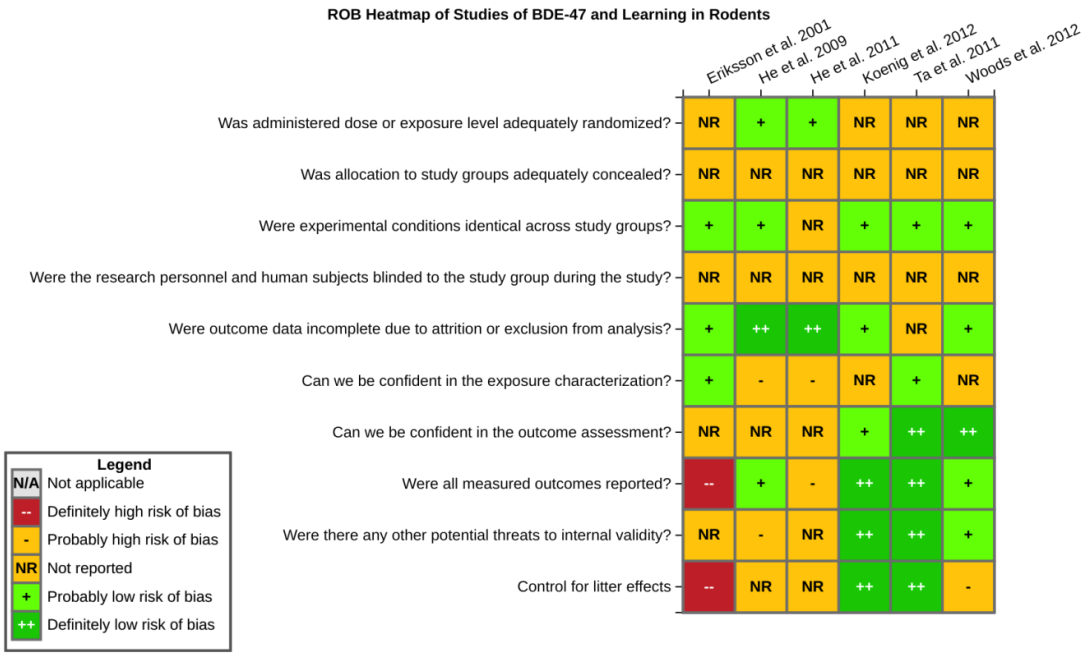


FIGURE 4-8 Risk of bias heatmap of studies of BDE-47 and learning in rodents. In HAWC: https://hawcproject.org/summary/visual/353/.

FIGURE 4-9 Risk of bias heatmap of studies of BDE-47 and memory in rodents. In HAWC: https://hawcproject.org/summary/visual/354/.

TABLE 4-3 Studies of BDE-47 and Memory in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
He et al. 2009	Sprague- Dawley rats	PND 10	2 months	Morris water maze	None	1
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
Koenig et al. 2012	C57BL/6J mice	GD 0 - PND 21	2 months	Barnes maze	1	None
Ta et al. 2011	C57BL/6J mice	GD 0 - PND 21	2 months	Morris water maze	1	None
Woods et al. 2012	Female Mecp2 308+/- mice	GD 0 - PND 21	PND 54	Morris water maze	None	0.03
	Male Mecp2 308+/- mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Female C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Male C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

BDE-99 and Learning and Memory

There is moderate confidence in the body of evidence on developmental exposure to BDE-99 and effects on learning and memory in rodents. Five studies of BDE-99 and learning were available (see Appendix E, Table E4-3). Two of the three studies in rats reported slower learning in the Morris water maze at a dose of 2 mg/kg-day (Cheng et al. 2009; Blanco et al. 2013). Zhao et al. (2014) reported no effects at a lower dose (0.2 mg/kg-day) under similar exposure and testing conditions. In contrast, developmental exposure of Wistar rats at doses up to 30 mg/kg-day had no effect on learning tested with a Y maze (Llansola et al. 2009). A single study (Fischer et al. 2008) in NMRI mice also reported decrements in learning during the acquisition period in tests using either a radial maze or a Morris water maze at a dose of 0.8 mg/kg-day. As noted earlier for BDE-47, the results of a meta-analysis (presented later in this chapter) of rodent studies on several BDEs, including BDE-99, and latency in the last trial of the Morris water maze lessened the committee's concerns about unexplained inconsistency. Three studies of BDE-99 and memory found no effects in several memory tests at doses of 0.2-2 mg/kg-day (see Appendix E, Table E4-4).

Confidence in the evidence on both learning and memory was downgraded because of serious concerns about several risk of bias issues. Risk of bias heatmaps of the studies are available in Appendix E (see Figures E4-3 and E4-4). The study with the fewest concerns in study design and conduct (Blanco et al. 2013) was rated probably high risk of bias for at least one key risk of bias issue (e.g., lack of randomization of treatment), and at least one study had a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis. No other downgrades or upgrades on other factors were made.

BDE-153 and Learning and Memory

There is low confidence in the body of evidence on developmental exposure to BDE-153 and effects on learning and memory. Two studies—one in rats (Zhang et al. 2013) and one in mice (Viberg et al. 2003)—evaluated both outcomes (see Appendix E, Tables E4-5 and E4-6). Although the results were inconsistent, a meta-analysis (presented later in this chapter) of data on several BDEs strengthened the evidence for an effect on learning. Confidence was downgraded twice for serious concerns about multiple risk of bias concerns, including lack of randomization of treatment, reduced confidence in outcome assessment due to lack of blinding of outcome assessors, and definitely high risk of bias ratings for not controlling for litter effects in the study design or analysis. A risk of bias heatmap of the studies is available in Appendix E (see Figure E4-5). No other downgrades or upgrades on other factors were made.

BDE-203 and Learning and Memory

There is very low confidence in the body of evidence on developmental exposure to BDE-203 and effects on learning and memory in mice. One study was available on both end points (Viberg et al. 2006), and only a single dose was tested in a single species (mouse). Therefore, confidence was downgraded for both outcomes because it was not possible to evaluate consistency in results (see Appendix E, Tables E4-7 and E4-8). The study did not have elements that would strengthen conclusions from a single study, such as multiple species, strains, or a particularly large sample size. Confidence was also downgraded twice because of multiple risk of bias issues, including lack of randomization of treatment, reduced confidence in outcome assessment due to lack of blinding of outcome assessors, and a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis. A risk of bias heatmap of the study is available in Appendix E (see Figure E4-6). No other downgrades or upgrades on other factors were made.

BDE-206 and Learning

There is very low confidence in the body of evidence on developmental exposure to BDE-206 and learning in mice. Only one study was available (see Appendix E, Table E4-9). Confidence in the body of evidence was downgraded because only a single study in mice was identified (Viberg et al. 2006), and it was not possible to establish or evaluate consistency as the study tested just a single species and a single dose. The study did not have elements that would strengthen conclusions from a single study, such as multiple species, strains, or a particularly large sample size. Confidence was also downgraded twice because of multiple risk of bias issues, including lack of randomization of treatment, reduced confidence in outcome assessment due to lack of blinding of outcome assessors, and a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis. A risk of bias heatmap of the study is available in Appendix E (see Figure E4-6). No other downgrades or upgrades on other factors were made.

BDE-209 and Learning and Memory

There is moderate confidence in the body of evidence on developmental exposure to BDE-209 and effects on learning and low confidence in the body of evidence on memory. Eight studies on learning were available (see Appendix E, Table E4-10). Several studies show effects on learning when it was assessed using the Morris water maze; however, other studies found no effects when BDE-209 was tested at similar doses using other test methods. As noted earlier for BDE-47, a meta-analysis (presented later in this chapter) of data on several BDEs, including BDE-209, strengthened the evidence for an effect on learning. Six studies on memory were available on BDE-209 (see Appendix E, Table E4-11). Several mouse studies showed effects on memory when assessed with the Morris water maze, but other studies found no effects on memory in the same dose range using other methods. Thus, confidence in the memory evidence was downgraded for unexplained inconsistency. Confidence in the evidence on both outcomes was downgraded because of concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors and a definitely high risk of bias rating in three studies because of failure to control for litter effects in the study design or analysis. Risk of bias heatmaps of the studies are available in Appendix E (see Figure E4-7 and E4-8).

DE-71 and Learning, Memory, and Attention

There is very low confidence in the body of evidence to evaluate whether developmental exposure to DE-71 affects learning, memory, or attention in rats. There were three studies on learning, two on memory, and three on attention (see Appendix E, Tables E4-12, E4-13, and E4-14). The results of the three studies on learning were inconsistent and used different tests (Morris water maze, radial maze, and visual discrimination) and animals of different ages. One study (Dufault et al. 2005) reported increased

errors in the visual discrimination task at the single dose tested (30 mg/kg-day on postnatal days 6-12). The other two studies reported no effects of DE-71 on learning at the same dose using longer exposure windows (Bowers et al. 2015; de-Miranda et al. 2016); however, the animals in these studies were evaluated at older ages than were the rats tested in the Dufault et al. (2005) study. The two studies of DE-71 effects on memory had inconsistent results and were evaluated in rats of different ages and with different tests (Morris water maze and radial maze). One study (de-Miranda et al. 2016) reported memory deficits in female Wistar rats (but not in males) in the radial maze at the single dose tested (30 mg/kg-day). Studies of DE-71 and attention were from a single laboratory (Dufault et al. 2005; Driscoll et al. 2009, 2012), and the majority of the tests reported no effects at doses up to 30 mg/kg-day across multiple tests (various attention tasks and a visual task).

Confidence in the body of evidence for all three outcomes was downgraded for unexplained inconsistency and downgraded twice for serious concerns about multiple risk of bias issues, such as reduced confidence in outcome assessment due to lack of reporting about whether outcome assessors were blinded and definitely high risk of bias ratings for exposure characterization in two of the studies (Dufault et al. 2005; de-Miranda et al. 2016). Risk of bias heatmaps of the studies are available in Appendix E (see Figures E4-9, E4-10, and E4-11). No downgrades or upgrades on other factors were made.

A summary of the confidence ratings of all the BDEs is presented in an evidence profile in Table 4-4.

Meta-Analysis of Selected Animal Data on Learning

The animal database for effects from PBDEs is both diverse and complex, with studies of varying designs and varying outcome measures. The outcome judged to be most amenable to meta-analysis was the results for latency in the last trial of the Morris water maze. This maze was the test used most often in the PBDE studies. It is a test of spatial learning for rodents that requires them to use distal cues to navigate from starting locations around the perimeter of an open swimming area to locate a submerged platform. Learning is assessed by latency, the amount of time it takes the animal to find the platform across repeated trials; learning is demonstrated by a reduction in latency with an increasing number of trials. For the meta-analysis, latency data on the last trial were used because latency to find the platform on the last trial was always reported in these studies. There are no a priori data that suggest species differences, so results for rats and mice were analyzed together. Given the sparse data on individual PBDEs, all BDEs were initially analyzed together. Additional analyses of individual PBDEs were conducted in cases where there were more than two data points (see Appendix E, Section E-5). All studies were considered except one that used humanized transgenic mice (which had variants of a human APOE gene).

Effect sizes were calculated as the $\log_{10}$ ratio of the mean difference between the treatment group and the concurrent control, multiplied by 100 ($y = 100 \times \ln$ [mean of treated group ÷ mean of control group]). For small changes, this is approximately equal to the percent change, but the resulting confidence interval is more symmetric and closer to normal (Hedges et al. 1999; Lajeunesse 2011). This normalization allows for treatment groups to be compared across studies and experiments. When normalized in this way, however, treatment groups within a study are correlated. Therefore, in one of the sensitivity analyses, effects were estimated using only the highest treatment group from each study. Additional sensitivity analyses were performed by sequentially excluding each study (all treatment groups for that study). See Appendix E, Section E-5, for the sensitivity analyses.

Both the overall effect of any treatment and the coefficients of meta-regressions were estimated. For meta-regressions, three models were used: a linear model in $y = a + b*\log_{10}(\text{dose})$ in order to test for a dose-response trend; a linear model $y = b*\text{dose}$; and a linear-quadratic model $y = b*\text{dose} + c*\text{dose}^2$ in order to model the dose-response shape. In the linear and linear-quadratic models, the intercept was omitted because the effect measures were already normalized relative to control levels. Additionally for these models, the coefficients were rescaled in terms of the change per 10 mg/kg-day (e.g., $y = b*[\text{dose}/10] + c*[\text{dose}/10]^2$) for ease of interpretation. In all cases, random effect models were used, as described in the

TABLE 4-4 Profile of the Body of Evidence on PBDEs and Learning, Memory, and Attention

	Factors Decreasing Confidence "---" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "---" If Not Present; "↑" If Sufficient to Upgrade Confidence					
INITIAL CONFIDENCE for each body of evidence (# of studies)	**Risk of Bias**	**Unexplained Inconsistency**	**Indirectness**	**Imprecision**	**Publication Bias**	**Large Magnitude**	**Dose Response**	**Residual Confounding**	**Consistency Species/Model**	**FINAL CONFIDENCE RATING**	**LEVEL OF EVIDENCE FOR HEALTH EFFECT***
BDE-47 Learning High (2 rat,[a] 4 mouse[b])	↓	---	---	---	---	---	---	---	---	Moderate	Moderate
BDE-47 Memory High (1 rat,[c] 4 mouse[b])	↓	↓-	---	---	---	---	---	---	---	Low	Low
BDE-99 Learning High (4 rat,[d] 1 mouse[e])	↓	---	---	---	---	---	---	---	---	Moderate	Moderate
BDE-99 Memory High (2 rat,[f] 1 mouse[e])	↓	---	---	---	---	---	---	---	---	Moderate	Inadequate
BDE-153 Learning High (1 rat,[g] 1 mouse[h])	↓↓	---	---	---	---	---	---	---	---	Low	Low
BDE-153 Memory High (1 rat,[g] 1 mouse[h])	↓↓	---	---	---	---	---	---	---	---	Low	Low
BDE-203 Learning High (1 mouse[i])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate
BDE-203 Memory High (1 mouse[i])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate
BDE-206 Learning High (1 mouse[i])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate
BDE-209 Learning High (2 rat,[j] 6 mouse[k])	↓	---	---	---	---	---	---	---	---	Moderate	Moderate
BDE-209 Memory High (1 rat,[l] 5 mouse[m])	↓	↓-	---	---	---	---	---	---	---	Low	Low
DE-71 Learning High (3 rat[n])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate
DE-71 Memory High (2 rat[o])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate
DE-71 Attention High (3 rat[p])	↓↓	↓	---	---	---	---	---	---	---	Very Low	Inadequate

*See the section "Determinations of Level of Evidence" later in this chapter for an explanation of how these ratings were determined.

NOTE: Studies were available on six BDE congeners and one technical grade mixture. All the BDEs had studies of learning, six had studies of memory, and only the mixture had studies of attention; no studies of response inhibition were found for any of the congeners.

[a]He et al. (2009, 2011).
[b]Eriksson et al. (2001); Ta et al. (2011); Koenig et al. (2012); Woods et al. (2012).
[c]He et al. (2009).
[d]Cheng et al. (2009); Llansola et al. (2009); Blanco et al. (2013); Zhao et al. (2014).
[e]Fischer et al. (2008).
[f]Blanco et al. (2013); Zhao et al. (2014).
[g]Zhang et al. (2013).
[h]Viberg et al. (2003).
[i]Viberg et al. (2006).
[j]Biesemeier et al. (2011); Y. Chen et al. (2014).
[k]Eriksson et al. (2001); Rice et al. (2009); Reverte et al. (2013); Verma et al. (2013, 2014); Buratovic et al. (2014).
[l]Biesemeier et al. (2011).
[m]Eriksson et al. (2001); Reverte et al. (2013); Verma et al. (2013, 2014); Buratovic et al. (2014).
[n]Dufault et al. (2005); Bowers et al. (2015); de-Miranda et al. (2016).
[o]Bowers et al. (2015); de-Miranda et al. (2016).
[p]Dufault et al. (2005); Driscoll et al. (2009, 2012).

protocol. All analyses utilized random effects models as implemented in the R "metafor" package. Sensitivity analyses included leaving one study out at a time and using only the highest dose group in each study (see Appendix E, Table E5-1). Benchmark dose (BMD) estimates were calculated for an effect size of 5% (BMD_5; see Appendix E, Figure E5-3). The BMD_5 was calculated using the linear or the linear-quadratic model, with the model selection based on the lowest AICc (Akaieke information criterion corrected for small sample size). The BMD_5 was calculated only for the "fixed effect"—that is, the estimated mean response across studies.

A meta-analysis using all the data is shown in Figure 4-10 and includes data from DE-71, BDE-47, -99, -153, and -209. For some studies, the standard deviations (SDs) were not reported or could not be digitized. These are shown in blue in the figure, and they were not included in the meta-analysis. There was a concern about possible reporting bias because studies that reported an SD might be more likely to show an effect than studies that did not. No such bias is evident from the figure, however, as the data from studies that did not have an SD had similar reported effect sizes as studies that did report an SD. For instance, the unweighted mean for studies reporting an SD was 30, whereas the unweighted mean for studies not reporting an SD was 36. Therefore, excluding studies without reported SDs is not likely to lead to a substantial bias in the meta-analysis results. The results of a meta-analysis of studies for which SDs were available is presented in Figure 4-11.

The results are as follows:

- Statistically significant overall effect of PBDE treatment that is also robust to leaving out individual studies, using only the highest dose group in each study and leaving out individual studies using the highest dose group only. There was also low or no heterogeneity (25%, not statistically significant in primary analysis; and <35%, not statistically significant in all sensitivity analyses).
- There was a positive, but not statistically significant, trend from the meta-regression in $\log_{10}$(dose). Meta-regression using a linear model or a linear-quadratic model resulted in a statistically significant linear term. The estimated BMD_5 for change in latency was 5.1 mg/kg-day (95% confidence interval [CI]: 3.2, 13) for the linear model and 1.8 mg/kg-day (95% CI: 1.1, 4.5) for the linear-quadratic model (see Appendix E, Figure E5-3). In both cases, however, heterogeneity was statistically significant with I^2 increased to >70%.

Overall, there is consistent evidence of an increase in latency in the last trial of the Morris water maze that is robust to multiple sensitivity analyses. Nevertheless, the evidence of a dose-response gradient across all PBDE congeners is tempered by the fact that there is no statistically significant trend in $\log_{10}$(dose) and that heterogeneity increased under linear and linear-quadratic meta-regression. When accounting for dose, this heterogeneity might be because different PBDEs have different potencies for this effect, different duration of dosing in the studies produces different cumulative doses, and test methods (e.g., number of daily water maze trials) vary among experiments. It is possible that using a different dose metric—such as cumulative dose or cumulative dose during a particular developmental window—would produce a dose-response gradient. Separate analyses of each PBDE for which there was enough data for meta-analysis were subsequently conducted (see Appendix E, Section E-5).

Human-Health Effects Results

Lam et al. (2015, 2016) conducted a systematic review of associations between developmental exposure to PBDEs and measures of intelligence and attention in children. Developmental exposure was defined as exposure that occurred prior to conception in one or both parents, during pregnancy (exposure to offspring in utero), perinatally, or in childhood. Studies were sought that measured exposure in human biological samples (e.g., urine, blood, or other specimens). The committee was provided with a draft of the systematic review in July 2016, and the committee reviewed it and decided to update it. In early 2017, the authors notified the committee that their review had been submitted for publication and that the litera-

ture search had been updated before the draft was submitted. The paper has been accepted for publication in *Environmental Health Perspectives* (Lam et al. in press). This section describes the committee's evaluations in the sequence they occurred, and includes a description of the draft systematic review by Lam et al., the committee's evaluation of it, the committee's update, and a discussion of the updated Lam et al. review that was accepted for publication.

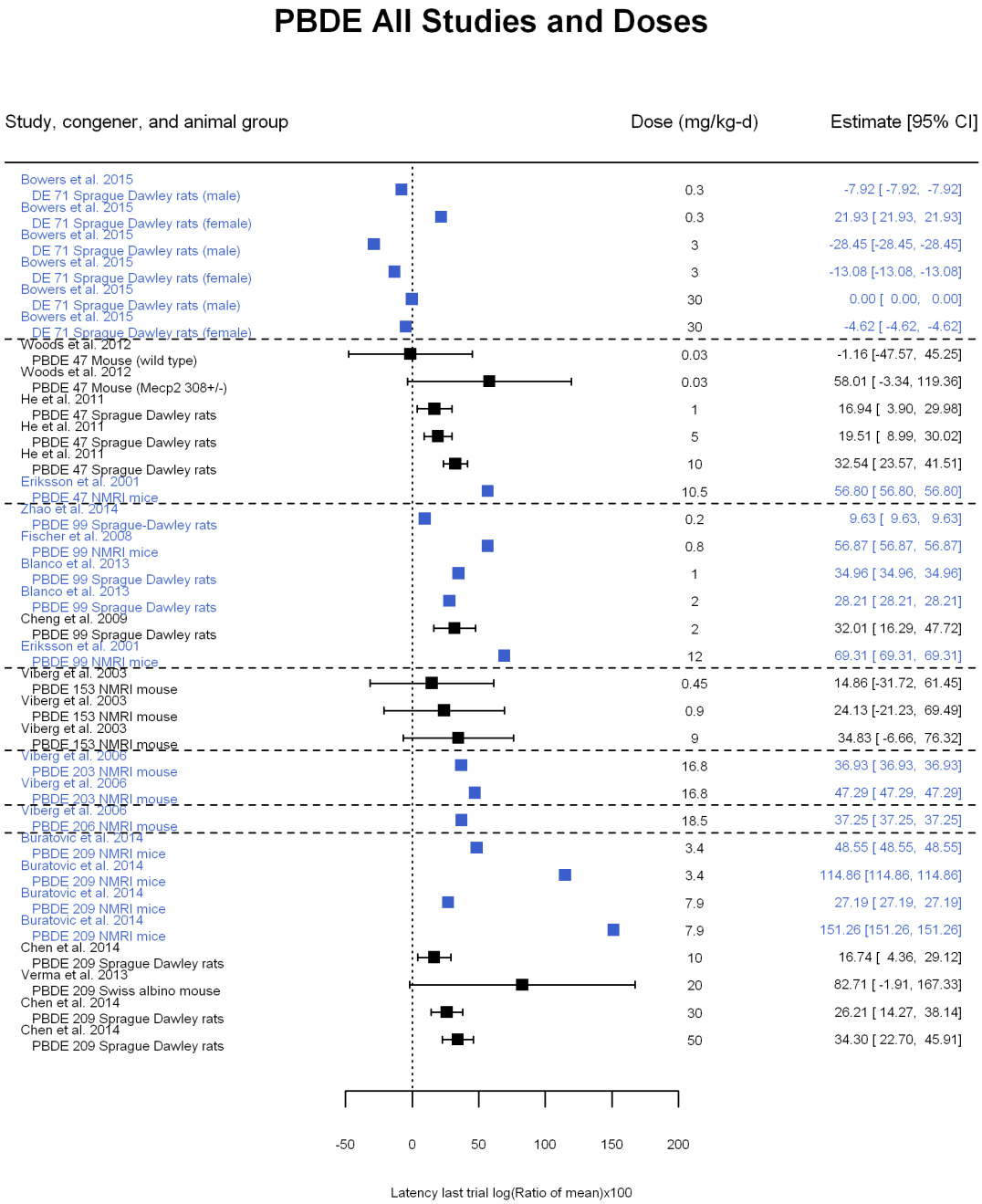


FIGURE 4-10 Forest plot of all studies of BDEs and latency in the last trial of the Morris water maze in rats and mice. Studies shown in black reported standard deviations while the studies in blue do not or the standard deviation could not be confidently digitized from the study figures. Dashed lines separate different congeners.

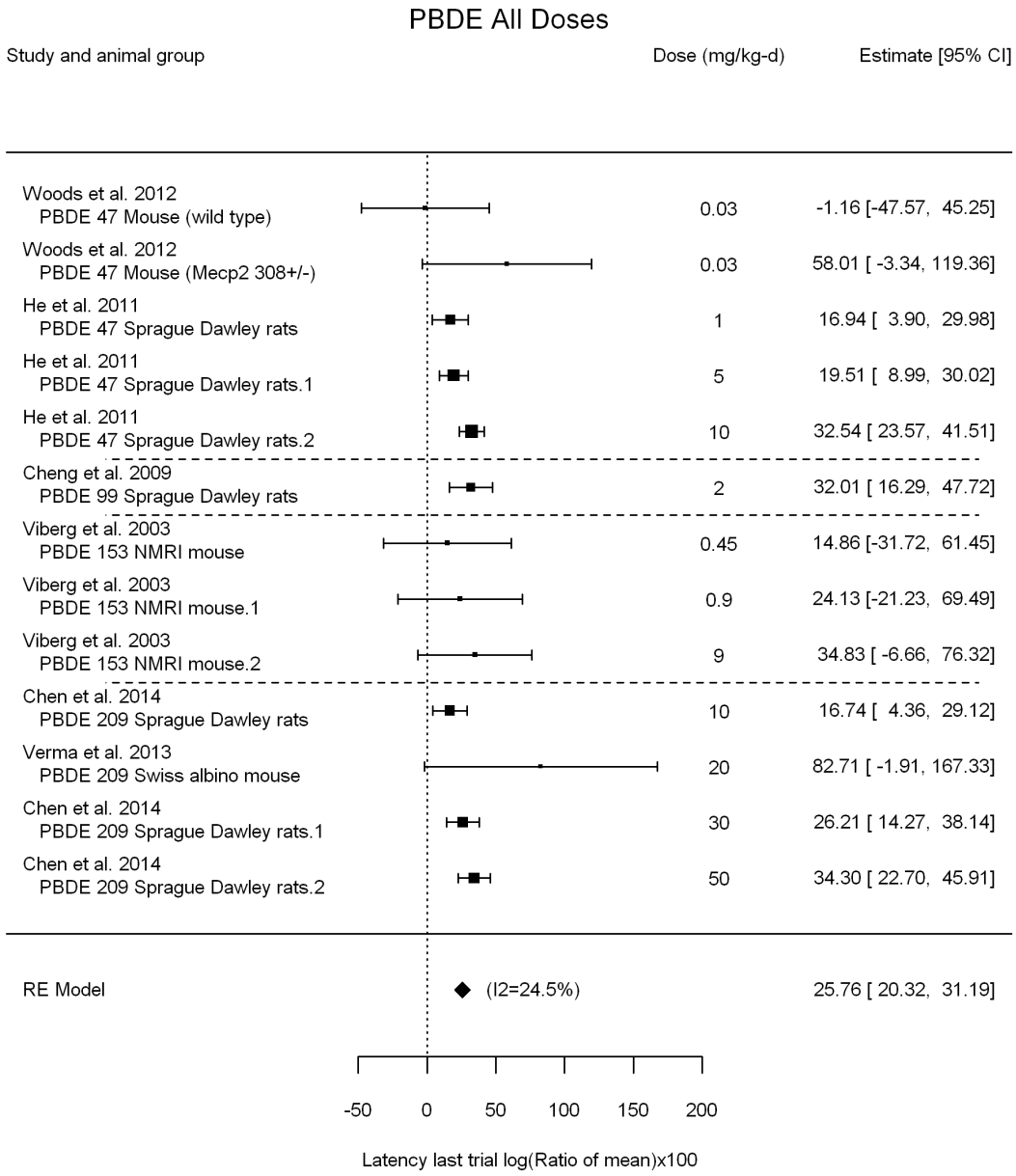


FIGURE 4-11 Results of the meta-analysis of PBDEs and latency in the last trial of the Morris water maze in rats and mice sorted by congener and then by dose. Analysis was restricted to studies for which standard deviations were reported or could be digitized from figures presented in the publications. Dashed lines separate different congeners. The overall effect of treatment is shown at the bottom of the figure as the change per 10 mg/kg-day.

As described earlier under "Evaluation of the Lam et al. (2015, 2016) Review," the draft systematic review was relevant to the committee's topic of interest and was judged appropriate for demonstrating how an existing systematic review could be updated by EPA. The original literature search was performed on March 5, 2015, without any date restrictions. After screening the results, 12 studies met the inclusion criteria of the PECO statement; nine studies measured IQ and seven studies evaluated ADHD and/or attention-related behavioral conditions (see Appendix F, Section F-4). Most of the individual IQ studies were judged to have low or probably low risk of bias, and the authors rated the confidence (or "quality") in the body of evidence as moderate. Only one study was given a high risk of bias rating in one of the domains (selective outcome reporting [Lin et al. 2010]); the study was part of conference proceedings and

did not provide sufficient detail about effects estimates for all the study outcomes. More concerns about risk of bias were found in the studies of ADHD and attention-related behaviors. Two studies were given high risk of bias ratings, one in the domain of confounding (Gump et al. 2014) and one in the domain of incomplete outcome data (Roze et al. 2009); the latter study was also given probably high risk of bias rating in two domains because of lack of blinding and inadequate adjustment for confounding.

A meta-analysis including four of the nine IQ studies (Herbstman et al. 2010; Gascon et al. 2012; Eskenazi et al. 2013; A. Chen et al. 2014) was performed and found a decrease in IQ in relation to PBDE exposure. Under the Navigation Guide, the confidence (or "quality") of the body of evidence was rated as moderate and the strength of evidence was considered sufficient[1] by the authors and was translated by the committee to a "moderate" level of evidence to support an inverse association between PBDEs and IQ following the OHAT method.

Although two of the individual studies that evaluated ADHD and attention-related behavioral conditions had high risk of bias rating in at least one domain, the other five studies were found to have low or probably low risk of bias, and the authors rated the confidence of the body of evidence as moderate. [2] They also judged that there was an insufficient number of combinable studies to perform a meta-analysis. Under the Navigation Guide, the strength of evidence was considered limited by the authors and was translated by the committee to a "low" level of evidence to support an association between PBDEs and ADHD following the OHAT method.

The committee's search for articles published since the Lam et al. (2015, 2016) literature search found three articles that met the eligibility criteria. All three articles involved cohorts from studies included in the Lam et al. systematic review (see Appendix F, Section F-4). One of the reports (Zhang et al. 2017) assessed full-scale IQ scores at age 9 in the same cohort as Y. Chen et al. (2014). The IQ data from this cohort at earlier ages were included in the meta-analysis of IQ scores. The effect size (a decrease of 5.3 IQ points for each 10-fold increase in PBDE exposure) reported by Zhang et al. (2017) was consistent with the earlier IQ data from the cohort and was very similar in magnitude to the overall effect size in the meta-analysis. The committee concluded that a new meta-analysis would not change the strength of evidence, nor would it alter the overall conclusion reached in the Lam et al. (2015, 2016) systematic review of an inverse association between PBDEs and childhood IQ.

Each of the three newly identified articles assessed attention-related problems and/or ADHD (see Appendix F, Section F-4). Zhang et al. (2017) found that each 10-fold increase in serum PBDE concentration was marginally associated with a 3.5-point increase in externalizing problems scores on the BASC-2 (Behavioral Assessment System for Children-2), and that was consistent with the results of earlier tests conducted in the cohort (Y. Chen et al. 2014). The article by Cowell et al. (2015) reported the results of behavioral assessments of the same cohort of children studied by Herbstman et al. (2010). The CBCL (Child Behavior Checklist) was administered annually from age 3 through age 7. Multivariable regression analyses were performed on the data collected at ages 4 and 6 because they were the oldest ages at which the preschool and school-aged CBCLs were performed in person rather than over the phone. Associations were detected between cord blood concentration of BDE-47 and BDE-153 and increased attention problems in children at age 4 but not age 6. Sagiv et al. (2015) conducted assessments of attention and ADHD in children from the same cohort studied by Eskenazi et al. (2013). Measures included the Conners' Con-

[1]Lam et al. based their conclusion for IQ on (1) moderate quality/confidence in the body of evidence for IQ, (2) consistent evidence for an effect from BDE-47, other congeners, and overall consistent results in combination of similar studies in a meta-analysis; and (3) support from one or more well-conducted studies; most studies were prospective cohorts that as a group represented diverse populations, were reasonably large, and supported by a statistically significant meta-analysis. The Navigation Guide has additional steps to reach a "strength of evidence" conclusion. There is no equivalent step in the OHAT method.

[2]Lam et al. based their conclusion for attention-related behaviors on (1) moderate quality/confidence in the body of evidence for attention and (2) general evidence for an effect from BDE-47 and other congeners, but not consistent evidence overall. There were too few combinable studies to conduct a meta-analysis, so chance, bias, and confounding could not be ruled out with reasonable confidence. The Navigation Guide has additional steps to reach a "strength of evidence" conclusion. There is no equivalent step in the OHAT method.

tinuous Performance Test II (CPT II), the ADHD Confidence Index score, which is derived from the Conners' CPT II, and the Conners' ADHD Index (CADS-P), which is derived from the Conners' Parent Rating Scale. They reported associations of higher prenatal serum concentrations of PBDEs with decrements in attention on the Conners' CPT II task as well as with increased ADHD Index scores at both 9 and 12 years of age; they also reported an association of higher prenatal PBDE exposure with increased scores on the CADS-P at age 9 but not at age 12. The results were consistent with the earlier findings in this cohort.

The committee concluded that, even with the addition of the three new reports, the data are still not amenable to meta-analysis due to the different assessment measures used across studies that prevent them from being combined. Furthermore, although all three reports found associations between prenatal serum PBDE concentrations and at least one measure of attention-related problems or ADHD, these reports present data collected from study cohorts that were already included in the Lam et al. review. The committee judged that they would not appreciably change the strength of evidence, nor would they change the overall conclusions reached by Lam et al. (2015, 2016) of limited evidence to support an association between PBDEs and ADHD.

After the committee completed its analysis, Lam et al.'s systematic review was accepted for publication (Lam et al. in press). The literature search had been updated from the 2016 draft, and the same three new publications identified as relevant to the systematic review by the committee were also found by the authors and included in their updated systematic review. The inclusion of the new evidence did not change the risk of bias assessments, but an analysis of the ratings in relation to age found that most studies that tested children at a later age (tested with Full Scale IQ) were rated as having low or probably low risk of bias across domains, whereas many studies of children at younger ages (tested with the Bayley Scales of Infant Development) had a rating of probably high risk of bias in at least one domain. The studies included in the meta-analysis remained unchanged, and the overall decrement in IQ points was reported as 3.70 (95% CI: 0.83, 6.56) per 10-fold increase in lipid-adjusted PBDE concentration (range: limit of detection–761 ng/g lipid). The committee's translation of the Navigation Guide evaluation of the into OHAT ratings remains the same—a moderate level of evidence to support an inverse association between PBDEs and IQ and a low level of evidence to support an association between PBDEs and ADHD.

MECHANISTIC EVIDENCE

The mechanisms of action through which developmental exposure to PBDEs alters neurobehavioral outcomes, such as IQ or attention in children or learning and memory in rodents, are not well understood. Nevertheless, data from mechanistic studies conducted in vitro or in animal models can help establish the biological plausibility of the associations that have been observed between PBDE exposure during the perinatal period and later behavioral outcomes. A large number of molecular, cellular, hormonal, and neurochemical changes have been reported following PBDE exposure (e.g., see Dingemans et al. 2011; Westerink 2014). In vitro, zebrafish and rodent models have been employed. An example of a possible initiating event is thyroid hormone disruption (Ibhazehiebo et al. 2011). Thyroid hormone plays a number of critical roles in brain development (Horn and Heuer 2010), and inadequate concentrations of thyroid hormone during early development have been associated with neurodevelopmental sequelae, including reduced IQ (Ghassabian et al. 2014), and increased risk of ADHD behaviors (Modesto et al. 2015). PBDEs have also been shown to alter intracellular calcium signaling through both the ryanodine and the IP3 receptors (Kim et al. 2011; Gassmann et al. 2014), leading to increased cytosolic calcium concentrations. PBDEs have been shown to increase oxidative stress in neuronal cell cultures (Costa et al. 2015), leading to apoptosis; to alter the expression of various genes involved in neurogenesis (Dingemans et al. 2011); and to decrease the expression of key proteins involved in neurodevelopmental processes (Kodavanti et al. 2015).

Examples of the wide array of potential initiating events are given in Figure 4-12. Although the figure is not exhaustive, it illustrates that there are many possible pathways through which developmental PBDE exposure could affect later cognitive or behavioral function. The large number of alterations that

have been reported at various levels, from molecular to neural systems, makes defining a particular adverse outcome pathway or pathways very difficult. Most studies span only one or at most two levels in the pathway. For example, Costa et al. (2015) have shown that BDE-47 causes oxidative stress leading to apoptosis both in cerebellar granule cells in vitro and in a mouse model. These changes were observed in the absence of any changes in thyroid hormone concentrations. Nevertheless, whether the observed increases in oxidative stress and subsequent apoptosis lead to changes in nervous system connectivity and behavior has not been investigated.

The development of high-throughput approaches to assess the effects of toxicants on the developing nervous system—especially those that may affect learning and memory—is a tremendous challenge. It is clear that many critical cell and molecular processes are involved. The adverse outcome pathways (AOPs) that have been proposed to lead to developmental neurotoxicity not only clearly identify major knowledge gaps in terms of the key events that are involved but also provide a valuable means to organize essential information and to identify research gaps (Bal-Price et al. 2017). Current approaches to fill some of these gaps include platforms based on the use of stem-cell-derived neurons or neuroprogenitor cells (Druwe et al. 2015; Pallocca et al. 2016; Ryan et al. 2016; Singh et al. 2016; Schmidt et al. 2017). End points include the expression of genes that play a role in neurodevelopment, neuroprogenitor cell proliferation and differentiation, neurite outgrowth, cell migration, and apoptosis as well as more generalized stress responses. Because the nervous system, ultimately, must develop a functional, coordinated, neural network, efforts are being expended to capture neural networks on microelectrode arrays (Brown et al. 2016). The usefulness of model species, such as *Caenorhabditis elegans* and zebrafish, is also being explored (Behl et al. 2015).

EVIDENCE INTEGRATION

Evidence synthesis was conducted in a three-part process. First, the confidence ratings for the animal studies were translated into conclusions about level of evidence of health effects using the procedure in Figure 4-4, and the level of evidence of health effects for the human studies was derived from the Lam et al. (2016, in press) analysis of the evidence. Second, initial hazard identification conclusions were reached by integrating the conclusions about level of evidence for the human and the animal evidence streams using the approach presented in Figure 4-5. Third, the degree of support from mechanistic data was considered and discussed in reaching final hazard identification conclusions.

Determinations of Level of Evidence

In the following sections, the confidence ratings of the evidence on each outcome and BDE congener are considered in context with the direction of the effect and then translated into a determination of level of evidence using Figure 4-4.

Animal Evidence

Learning

There is moderate confidence in the body of evidence on PBDEs and effects on learning in rodents on the basis of studies on BDE-47, -99, and -209. Qualitative analyses suggested effects on learning for each of these congeners. A meta-analysis of learning data found consistent evidence of an effect, measured as latency in the last trial of the Morris water maze. Thus, the moderate confidence rating translates to a moderate level of evidence that developmental exposure to these congeners is associated with decrements in learning in rodents.

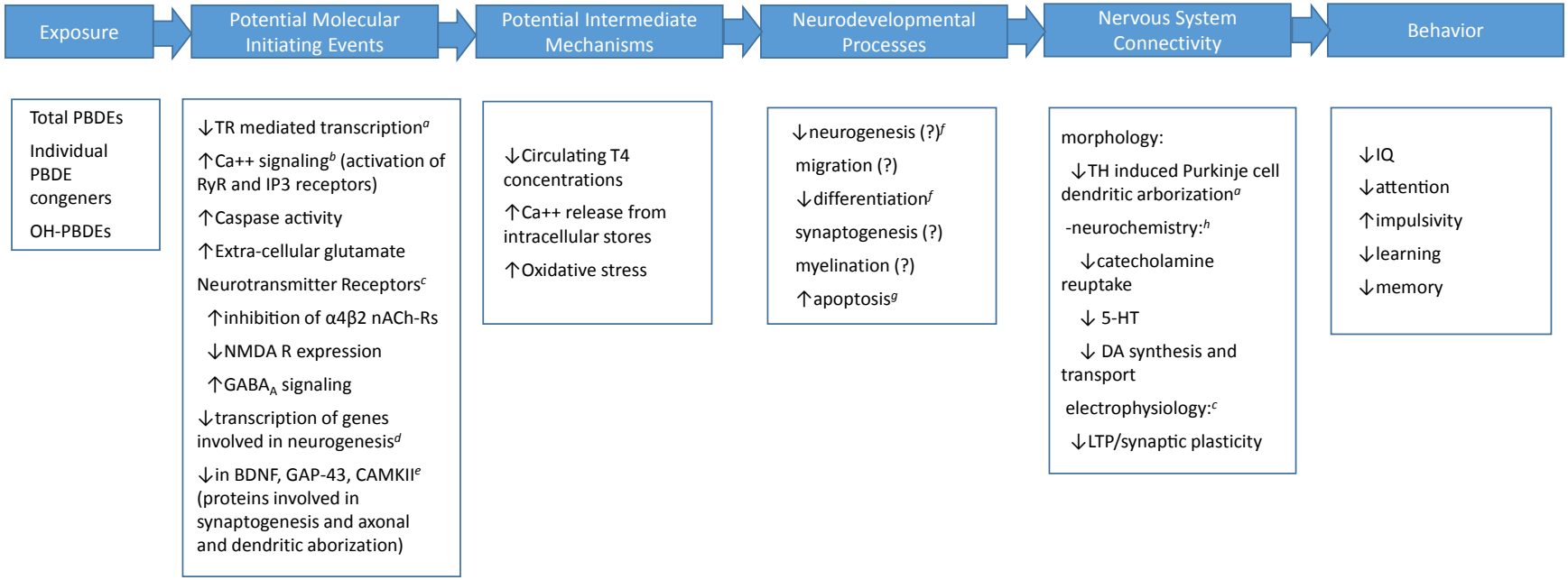


FIGURE 4-12 Theoretical steps involved in PBDE developmental neurotoxicity.
NOTE: 5-HT, 5-hydroxytryptamine (serotonin); BDNF, brain-derived neurotrophic factor; CAMKII, calmodulin-dependent kinase II; DA, dopamine; GABA, gamma-aminobutyric acid; GAP-43, growth-associated protein; IP3, 1,4,5-triphosphate; LTP, long-term potentiation; nACh-R, nicotine acetylcholine; NMDAR, glutamate receptor; RyR, ryanodine receptor; T4, thyroxine; TH, thyroid hormone; TR, thyroid hormone receptor.
[a]Ibhazehiebo et al. (2011).
[b]Kim et al. (2011); Gassmann et al. (2014); Westerink (2014).
[c]Westerink (2014).
[d]Wang et al. (2016).
[e]Viberg et al. (2008).
[f]Li et al. (2013).
[g]Costa et al. (2015).
[h]Wang et al. (2015, 2016).
SOURCE: Adapted from Mundy (2016).

There was low confidence in the body of evidence on BDE-153. Only two studies were available, one in rats and one in mice, with differing results. Nevertheless, the meta-analysis strengthened support for an effect on learning, so the low confidence rating was judged to have a low level of evidence (rather than an inadequate level) that developmental exposure to BDE-153 is associated with decrements in learning in rodents.

Very low confidence ratings were given to BDE-203 and -206 and to DE-71, which means there is an inadequate level of evidence to assess whether exposure to these congeners or to the technical mixture is associated with decrements in learning in rodents.

Memory

There is low confidence in the body of evidence on developmental exposure to PBDEs and effects on memory in rodents on the basis of evidence on BDE-47, -153, and -209. The two studies on BDE-153 reported results that were consistent in direction (both decreased performance on a memory test) across species. The results of the studies on BDE-47 and BDE-209 were less consistent, but effects on memory were found in some studies. Thus, the low confidence in the body of evidence was translated to a low level of evidence for these congeners.

DE-71 and BDE-203 had very low confidence ratings, which means there is an inadequate level of evidence to assess whether exposure to them is associated with decrements in learning in rodents.

Confidence in the body of evidence on BDE-99 was rated as being moderate, but the findings in those studies suggested a lack of association. A more robust database is needed to support a finding of a lack of effect (NTP 2015), so the evidence was judged to be inadequate to reach a conclusion that developmental exposure to BDE-99 has no effect on memory.

Attention

The only available data on attention and PBDEs was on DE-71. Confidence in the body of evidence was rated as being very low, which means there is an inadequate level of evidence to assess whether exposure to this technical mixture is associated with effects on attention in rats.

Human Evidence

Intelligence

As described earlier, the evaluation by Lam et al. (2016, in press) was translated into an OHAT level of evidence determination of a moderate level of evidence that exposure to PBDEs is associated with a decrease in IQ.

ADHD/Attention-Related Behavioral Conditions

As described earlier, the evaluation by Lam et al. (2016, in press) was translated into an OHAT level of evidence determination of a low level of evidence that exposure to PBDEs is associated with increased reporting of ADHD symptoms.

Hazard Identification Conclusions

The animal evidence on learning and memory was considered to have the closest parallels to intelligence measured in human studies, and the animal evidence on attention was used in parallel with the human evidence on ADHD. Using the OHAT scheme presented in Figure 4-5, the hazard conclusions drawn

were that (1) developmental exposure to PBDEs is presumed to pose a hazard to intelligence in humans, and (2) it is not possible to draw a conclusion about potential hazards to attention-related behavioral conditions in humans. Because the mechanisms of action involved in developmental neurotoxicity are unknown, this data stream had minimal impact on the hazard identification conclusion.

ANALYSIS OF LOW-DOSE EFFECTS

Human studies provide a moderate level of evidence that PBDEs are associated with decrements in IQ in humans. Lam et al. (in press) found a decrease of 3.70 IQ points in children per 10-fold increase in serum PBDE concentration. The committee's update of the systematic review likewise did not find evidence that would appreciably change the overall conclusions reached by Lam et al. (2016) of limited evidence (or a low level of evidence according to the OHAT method) to support an association between PBDEs and ADHD.

The committee attempted to compare exposures to PBDEs in humans with those in animal studies (see Table 4-5). Human intakes of PBDEs estimated from levels found in food and dust appear to be about 500,000 times lower than the intakes estimated from animal studies (BMD_5 estimated earlier in the chapter). To compare internal doses, blood concentrations of BDE-47 in humans were obtained from the National Health and Nutrition Examination Survey (NHANES) and from one epidemiology cohort included in the systematic review of the human evidence (CHAMACOS cohort). Blood concentrations of BDE-47 in rodents were obtained from studies in which the animals were exposed near the BMD_5. Comparisons of internal doses of BDE-47 also showed large disparities in the exposure between humans and animals, but not as large as suggested by the intake data. For lipid adjusted BDE-47 plasma levels, the 95th percentile exposure in humans was about 170 times lower than the level in rats treated near the BMD_5; the highest BDE-47 level in any child in the CHAMACOS cohort was 36 times lower than that rat plasma level. Another observation was that internal dose measurements were more similar between rodents and humans on a ng/g serum basis than on a ng/g lipid basis. Thus, there is significant uncertainty associated with the internal dose correspondence between rodents and humans, and the most relevant dose metric is unclear. These observations indicate that it would be helpful to have internal dose measures in rodent studies, preferably measuring the same matrix and metabolites as in human biomonitoring studies to reduce uncertainty in rodent to human extrapolation.

RELEVANCE TO ANIMAL TOXICITY TESTING

Rodent and human outcome measures in the PBDE case studies were less parallel than were those in the phthalate case studies in Chapter 3, where AGD, alterations in testosterone, and hypospadias were shared outcome measures in the studies. In general, measurement of IQ in children relies on well-described test methods that have been validated for broad use in biomedical research. The committee's evaluation of the animal data found that, unlike the case of IQ testing in people, standardized test batteries, animal models, and exposure regimens were not used in the animal studies. Even within individual tests, different methods for categorizing responses were sometimes used, making it difficult to determine whether different studies shared similar or conflicting results or if they assessed different cognitive functions. Despite these limitations, the committee was able to demonstrate the use of a meta-analysis that combined results from a single type of test and a single outcome measure for several BDEs that showed a statistically significant overall effect of BDE treatment. This particular example represents a case where current toxicity-testing paradigms detect a hazard that is presumed to be of concern to humans but might not be not accurately predicting doses at which effects occur in humans.

TABLE 4-5 Comparison of Human and Rat Intake and Internal Concentrations of BDE-47

Human Intake or Blood Concentrations of BDE-47		Rat Administered Dose, Blood BDE Concentration, and BMD_5	
Description	Value	Description	Value
Total PBDE and BDE-47 estimated intake based on levels reported in food and dust[a]	mg/kg-day Total PBDEs: 7.7×10^{-6} (adults) 4.9×10^{-5} (ages 1-5) 1.4×10^{-5} (ages 6-11) 9.1×10^{-6} (ages 12-19) BDE-47: 2.6×10^{-6} mg/kg-day (adults)	PBDE BMD_5 for latency in the last trial of the Morris water maze[b]	1.79-5.08 mg/kg-day
BDE-47 in blood (NHANES 2003-2004)[c]	ng/g serum 0.12 (median) 1.0 (95th percentile) 13 (maximum)	BDE-47 measured levels in whole blood of mouse dam treated with 1 mg/kg-day for about 50 days before and during pregnancy[d]	28 ng/g blood
		BDE-47 measured levels in whole blood of mouse dam treated at 1 mg/kg-day for about 70 days before and during pregnancy and weaning[e]	9.6 ng/g blood
	ng/g lipid 19 (median) 163 (95th percentile) 2,350 (maximum)	BDE-47 measured levels in plasma of rats treated at 1 mg/kg-day for 14 days by gavage[f]	28,000 ng/g lipid
BDE-47 in serum at age 7 (CHAMACOS cohort)[g]	ng/g lipid 47.5 (geometric mean) 768 (maximum)		

[a]Lorber (2008).
[b]See Appendix E, Figure E5-3.
[c]NCHS (2007).
[d]Koenig et al. (2012).
[e]Ta et al. (2011).
[f]Darnerud et al. (2007).
[g]Bradman et al. (2012).

FINDINGS AND RECOMMENDATIONS

Systematic Reviews

- **Consistency and Transparency:** The committee found that the systematic review process was valuable because it provided a framework for identifying, selecting, and evaluating evidence in a consistent and explicit manner; maximized transparency in how the assessments were performed; and facilitated the clear presentation of the basis for scientific judgments.
- **Meta-analysis:** The committee found that the meta-analyses were valuable in summarizing data from the systematic reviews and in comparing the animal and human evidence in a robust and consistent manner. The meta-analysis of a subset of animal studies that tested learning in rodents exposed to various BDEs provided evidence of a possible relationship between PBDEs and decrements in learning that was not evident when the data sets on the individual BDEs were evaluated qualitatively. The meta-analysis results informed the confidence ratings of the body of evidence and allowed the committee to estimate benchmark doses on the basis of data from multiple studies. Meta-analyses of animal studies have not been commonly performed, but they were found to be useful both to inform confidence ratings in the body of evidence and to support benchmark dose modeling.

Recommendation: Systematic reviews should include meta-analyses of the animal and the human evidence, if approprriate. The results of meta-analyses should be used to examine quantitative relationships between EACs and end points of interest, to inform the confidence ratings of the bodies of evidence, and, if possible, to estimate benchmark doses.

- **Risk of Bias Evaluations:** Information important to the evaluation of the quality of individual animal studies was often not reported, including whether the study controlled for litter effects, whether animals were randomly allocated to study groups, and whether research personnel were blinded to the study groups during the outcome assessment. Because a lack of adequate reporting could not be distinguished from failure to adhere to practices that minimize bias, failure to report practices that minimize bias often led to higher risk of bias ratings for individual studies, downgrading the overall level of confidence in the body of evidence. These types of problems could be remedied if journals required better reporting of the methods used in animal studies, especially reporting pertaining to issues that might introduce bias into the research. These requirements could build on reporting standards that have been developed by various organizations to improve transparency (e.g., the ARRIVE guidelines [Kilkenny et al. 2010]). For example, studies should be required to report whether animals were assigned to study groups using random allocation and whether researchers were blinded to the study groups during outcome assessment.

- **Using an Existing Systematic Review:** The committee critically evaluated a recent systematic review of epidemiologic studies on the effect of developmental exposure to PBDEs on IQ and ADHD and judged it to be adequate for use in the context of the review question. This evaluation allowed the committee to focus its efforts on updating the literature search of the recent review. Because no studies on new cohorts were found, only a qualitative update to the recent review was performed.

  Recommendation: The US Environmental Protection Agency (EPA) should develop policies and procedures to allow the agency to use and update existing systematic reviews. It is important that the existing systematic review's study question directly addresses EPA's topic of interest and that the methods be critically evaluated before the systematic review is used and updated.

Evidence Integration

- A comparison of doses between the animal and the human studies of PBDEs was challenging and imprecise because the animal studies often report report external administered doses (usually without measures of internal dose), whereas human studies measure biomarkers of internal dose (with estimates of the external administered dose being uncertain). There is some indication that the difference in internal dose between humans and rodents may be less than the difference in administered dose; these estimates are uncertain, however, and additional work is needed to clarify this issue. Toxicology studies that measure internal dose metrics, including metrics that are similar to those used in human biomonitoring, could help address this data gap.

  Recommendation: To support animal-to-human extrapolations, pharmacokinetic data should be generated and used to develop pharmacokinetic models that make it possible to infer human internal doses (not just intake) from biomonitoring data and animal internal doses from administered doses.

- Integration of human data that evaluated measures of IQ and ADHD with animal studies that evaluated learning, memory, and attention was challenging since these end points are not identical. The animal studies use different tests of learning and memory and, even when the same type of behavioral test was used, testing methods and data analyses often differed between studies. The heterogeneity in testing methods and data analyses contributed to the challenges for evaluating consistency of the evidence in the animal studies. The committee found it helpful to focus its quantitative analysis on a specific measure of learning that was consistently reported in the animal studies.

Mechanistic Information

- A review of pharmacokinetic and mechanistic data on PBDEs in relation to developmental neurotoxicity provided some biological plausibility of the associations observed between PBDE exposure during the perinatal period and later neurobehavioral outcomes. An attempt to illustrate an adverse outcome pathway from these data was hindered by the complexity and multifactorial nature of how PBDEs affect neurodevelopmental processes.

Hazard Identification

- The committee concluded that developmental exposure to PBDEs is presumed to pose a hazard to intelligence in humans and that it was not possible to draw a hazard conclusion about effects on attention-related behavioral conditions in humans.

Low-Dose Effects

- The committee concluded that the human studies provide a moderate level of evidence that PBDEs are associated with decrements in IQ in humans. Uncertainty in the internal doses of humans relative to experimental animals limited the ability to draw conclusions about the prediction of low-dose effects based on experimental animal studies. The development of pharmacokinetic data and models for extrapolation of data from animal studies or human biomonitoring data could facilitate the evaluation of the potential of PBDEs to cause health effects in humans at low doses.

Other Considerations

- **Mixtures:** The committee found that humans are exposed to a mixture of PBDEs, whereas the experimental animal evidence was generally from studies with a single congener. This difference between the human mixture exposures and the single chemical animal exposures contributed to the challenges for integrating evidence between the human and the animal studies.
- **Expertise:** The committee found that the conduct of the systematic review and evidence integration requires a multidisciplinary approach that should be tailored to the specific review question. Experts in the conduct of meta-analyses and benchmark dose modeling will be essential.

REFERENCES

Bal-Price, A., P.J. Lein, K.P. Keil, S. Sethi, T. Shafer, M. Barenys, E. Fritsche, M. Sachana, and M.E. Meek. 2017 Developing and applying the adverse outcome pathway concept for understanding and predicting neurotoxicity. Neurotoxicology. 59:240-255.

Behl, M., J.H. Hsieh, T.J. Shafer, W.R. Mundy, J.R. Rice, W.A. Boyd, J.H. Freedman, E.S. Hunter, III, K.A. Jarema, S. Padilla, and R.R. Tice. 2015. Use of alternative assays to identify and prioritize organophosphorus flame retardants for potential developmental and neurotoxicity. Neurotoxicol. Teratol. 52(Pt. B):181-193.

Biesemeier, J.A., M.J. Beck, H. Silberberg, N.R. Myers, J.M. Ariano, A. Radovsky, L. Freshwater, D.W. Sved, S. Jacobi, D.G. Stump, M.L. Hardy, and T. Stedeford. 2011. An oral developmental neurotoxicity study of decabromodiphenyl ether (DecaBDE) in rats. Birth Defects Res. B Dev. Reprod. Toxicol. 92(1):17-35.

Blanco, J., M. Mulero, L. Heredia, A. Pujol, J.L. Domingo, and D.J. Sánchez**.** 2013. Perinatal exposure to BDE-99 causes learning disorders and decreases serum thyroid hormone levels and BDNF gene expression in hippocampus in rat offspring. Toxicology 308:122-128.

Bowers, W.J., P.M. Wall, J.S. Nakai, A. Yagminas, M. Wade, and N. Li. 2015. Behavioral and thyroid effects of in utero and lactational exposure of Sprague-Dawley rats to the polybrominated diphenyl ether mixture DE71. Neurotoxicol. Teratol. 52(Pt. B):127-142.

Bradman, A., R Castorina, A. Sjödin, L. Fenster, R.S. Jones, K.G. Harley, J. Chevrier, N.T. Holland, and B. Eskenazi. 2012. Factors associated with serum polybrominated diphenyl ether (PBDE) levels among school-age children in the CHAMACOS cohort. Environ. Sci. Technol. 46(13):7373-7381.

Bramwell, L., S.V. Glinianaia, J. Rankin, M. Rose, A. Fernandes, S. Harrad, and T. Pless-Mulolli. 2016. Associations between human exposure to polybrominated diphenyl ether flame retardants via diet and indoor dust, and internal dose: A systematic review. Environ. Int. 92-93:680-694.

Brown, J.P., D. Hall, C.L. Frank, K. Wallace, W.R. Mundy, and T.J. Shafer. 2016. Editor's highlight: Evaluation of a microelectrode array-based assay for neural network ontogeny using training set chemicals. Toxicol. Sci. 154(1):126-139.

Buratovic, S., H. Viberg, A. Fredriksson, and P. Eriksson. 2014. Developmental exposure to the polybrominated diphenyl ether PBDE 209: Neurobehavioural and neuroprotein analysis in adult male and female mice. Environ. Toxicol. Pharmacol. 38(2):570-585.

Butryn, D.M., M.S. Gross, L.H. Chi, A. Schecter, J.R. Olson, and D.S. Aga. 2015. "One-shot" analysis of polybrominated diphenyl ethers and their hydroxylated and methoxylated analogs in human breast milk and serum using gas chromatography-tandem mass spectrometry. Anal. Chim. Acta 892:140-147.

Canton, R., J. Sanderson, R. Letcher, A. Bergman, and M. van den Berg. 2005. Inhibition and induction of aromatase (CYP19) activity by brominated flame retardants in H295R human adrenocortical carcinoma cells. Toxicol. Sci. 88(2):447-455.

Canton, R.F., D.E. Scholten, G. Marsh, P.C. de Jong, and M. van den Berg. 2008. Inhibition of human placental aromatase activity by hydroxylated polybrominated diphenyl ethers (OH-PBDEs). Toxicol. Appl. Pharmacol. 227(1):68-75.

Ceccatelli, R., O. Faass, M. Schlumpf, and W. Lichtensteiger. 2006. Gene expression and estrogen sensitivity in rat uterus after developmental exposure to the polybrominated diphenylether PBDE 99 and PCB. Toxicology 220(2-3):104-116.

Chen, A., K. Yolton, S.A. Rauch, G.M. Webster, R. Hornung, A. Sjödin, K.N. Dietrich, and B.P. Lanphear. 2014. Prenatal polybrominated diphenyl ether exposures and neurodevelopment in U.S. children through 5 years of age: The HOME study. Environ. Health Perspect. 122(8):856-862.

Chen, L.J., E.H. Lebetkin, J.M. Sanders, and L.T. Burka. 2006. Metabolism and disposition of 2,2',4,4',5-pentabromodiphenyl ether (BDE99) following a single or repeated administration to rats or mice. Xenobiotica 36(6):515-534.

Chen, Y.H., Z.H. Li, Y. Tan, C.F. Zhang, J.S. Chen, F. He, Y.H. Yu, and D.J. Chen. 2014. Prenatal exposure to decabrominated diphenyl ether impairs learning ability by altering neural stem cell viability, apoptosis, and differentiation in rat hippocampus. Hum. Exp. Toxicol. doi: 10.1177/0960327113509661.

Cheng, J., J. Gu, J. Ma, X. Chen, M. Zhang, and W. Wang. 2009. Neurobehavioural effects, redox responses and tissue distribution in rat offspring developmental exposure to BDE-99. Chemosphere 75(7):963-968.

Cornell, J.E., C.D. Mulrow, R. Localio, C.B. Stack, A.R. Meibohm, E. Guallar, and S.N. Goodman. 2014. Randome-effects meta-analysis of inconsistent effects: A time for change. Ann. Intern. Med. 160(4):267-270.

Costa, L.G., C. Pellacani, K. Dao, T.J. Kavanagh, and P.J. Roque. 2015. The brominated flame retardant BDE-47 causes oxidative stress and apoptotic cell death in vitro and in vivo in mice. Neurotoxicology 48:68-76.

Cowell, W.J., S.A. Lederman, A. Sjödin, R. Jones, S. Wang, F.P. Perera, R. Wang, V.A. Rauh, and J.B. Herbstman**.** 2015. Prenatal exposure to polybrominated diphenyl ethers and child attention problems at 3-7 years. Neurotoxicol. Teratol. 52(Pt B):143-150.

Darnerud, P.O., and T. Sinjari. 1996. Effects of polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) on thyroxine and TSH blood levels in rats and mice. Organohalogen Compd. 29:316-319.

Darnerud, P.O., M. Aune, L. Larsson, and S. Hallgren. 2007. Plasma PBDE and thyroxine levels in rats exposed to Bromkal or BDE-47. Chemosphere 67(9):S386-S392.

de-Miranda, A.S., S.N. Kuriyama, C.S. da-Silva, M.S. do-Nascimento, T.E. Parente, and F.J. Paumgartten**.** 2016. Thyroid hormone disruption and cognitive impairment in rats exposed to PBDE during postnatal development. Reprod. Toxicol. 63:114-124.

Dingemans, M.M., M. van den Berg, and R.H. Westerink. 2011. Neurotoxicity of brominated flame retardants: (In)direct effects of parent and hydroxylated polybrominated diphenyl ethers on the (developing) nervous system. Environ. Health Perspect. 119(7):900-907.

Driscoll, L.L., A.M. Gibson, and A. Hieb. 2009. Chronic postnatal DE-71 exposure: Effects on learning, attention and thyroxine levels. Neurotoxicol. Teratol. 31(2):76-84.

Driscoll, L.L., J. Kaplan, E. Bucuvalas, H. Allen, J. Kraut, and J. Fitzpatrick. 2012. Acute postnatal exposure to the pentaBDE commercial mixture DE-71 at 5 or 15mg/kg/day does not produce learning or attention deficits in rats. Neurotoxicol. Teratol. 34(1):20-26.

Druwe, I., T.M. Freudenrich, K. Wallace, R.J. Shafer, and W.R. Mundy. 2015. Sensitivity of neuroprogenitor cells to chemical-induced apoptosis using a multiplexed assay suitable for high-throughput screening. Toxicology 333:14-24.

Dufault, C., G. Poles, and L.L. Driscoll. 2005. Brief postnatal PBDE exposure alters learning and the cholinergic modulation of attention in rats. Toxicol. Sci. 88(1):172-180.

EFSA (European Food Safety Authority). 2011. Scientific opinion on polybrominated diphenyl ethers (PBDEs) in food. EFSA J. 9(5):2156.

EPA (U.S. Environmental Protection Agency). 2010. An Exposure Assessment of Polybrominated Diphenyl Ethers. EPA/600/R-08/086F. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=210404 [accessed March 14, 2016].

Eriksson, P., E. Jakobsson, and A. Fredriksson. 2001. Brominated flame retardants: A novel class of developmental neurotoxicants in our environment? Environ. Health Perspect. 109(9):903-908.

Eskenazi, B., J. Chevrier, S.A. Rauch, K. Kogut, K.G. Harley, C. Johnson, C. Trujillo, A. Sjödin, and A. Bradman. 2013. In utero and childhood polybrominated diphenyl ether (PBDE) exposures and neurodevelopment in the CHAMACOS study. Environ. Health Perspect. 121(2):257-262.

Faass, O., R. Ceccatelli, M. Schlumpf, and W. Lichtensteiger. 2013. Developmental effects of perinatal exposure to PBDE and PBC on gene expression in sexually dimorphic rat brain regions and female sexual behavior. Gen. Comp. Endocrinol. 188:232-241.

Fischer, C., A. Fredriksson, and P. Eriksson. 2008. Coexposure of neonatal mice to a flame retardant PBDE 99 (2, 2',4, 4',5-pentabromodiphenyl ether) and methyl mercury enhances developmental neurotoxic defects. Toxicol. Sci. 101(2):275-285.

Fowles, J.R., A. Fairbrother, L. Baecher-Steppan, and N.I. Kerkvliet. 1994. Immunologic and endocrine effects of the flameretardant pentabromodiphenyl ether (DE-71) in C57BL/6J mice. Toxicology 86(1-2) :49-61.

Gascon, M., M. Fort, D. Martínez, A.E. Carsin, J. Forns, J.O. Grimalt, L. Santa Marina, N. Lertxundi, J. Sunyer, and M. Vrijheid. 2012. Polybrominated diphenyl ethers (PBDEs) in breast milk and neuropsychological development in infants. Environ. Health Perspect. 120(12):1760-1765.

Gassmann, K., T. Schreiber, M.M. Dingemans, G. Krause, C. Roderigo, S. Giersiefer, J. Schuwald, M. Moor, K. Unfried, K. Bergman, R.H. Weterink, C.R. Rose, and E. Fritsche. 2014. BDE-47 and 6-oh-BDE-47 modulate calcium homeostasis in primary fetal human neural progenitor cells via ryanodine receptor-independent mechanisms. Arch. Toxicol. 88(8):1537-1548.

Geyer, H.J., K.W. Schramm, P.O. Darnerud, M. Aune, E.A. Feicht, K.W. Fried, B. Henkelmann, D. Lenoir, P. Schmid, and T.A. McDonald. 2004. Terminal elimination half-lives of the brominated flame retardants TBBPA, HBCD, and lower brominated PBDEs in humans. Organohalogen Compd. 66:3820-3825.

Ghassabian, A., H. El Marroun, R.P. Peeters, V.W. Jaddoe, A. Hofman, F.C. Verhulst, H. Tiemeier, and T. White. 2014. Downstream effects of maternal hypothroxinemia in early pregnancy: Nonverbal IQ and brain morphology in school-age children. J. Clin. Endocrinol. Metab. 99(7):2383-2390.

Gump, B.B., S. Yun, and K. Kannan. 2014. Polybrominated diphenyl ether (PBDE) exposure in children: Possible associations with cardiovascular and psychological functions. Environ. Res. 132:244-250.

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

He, P., A.G. Wang, T. Xia, P. Gao, Q. Niu, L.J. Guo, and X.M. Chen. 2009. Mechanisms underlying the developmental neurotoxic effect of PBDE-47 and the enhanced toxicity associated with its combination with PCB153 in rats. NeuroToxicology 30(6):1088-1095.

He, P., A. Wang, Q. Niu, L. Guo, T. Xia, and X. Chen. 2011. Toxic effect of PBDE-47 on thyroid development, learning, and memory, and the interaction between PBDE-47 and PCB153 that enhances toxicity in rats. Toxicol. Ind. Health 27(3):279-288.

He, Y. M.B. Murphy, R.M. Yu, M.H. Lam, M. Hecker, J.P. Giesy, R.S. Wu, and P.K. Lam. 2008. Effects of 20 PBDE metabolites on steroidogenesis in the H295R cell line. Toxicol. Lett. 176(3):230-238.

Hedges, L.V., J. Gurevitch, and P.S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80(4):1150-1156.

Herbstman, J.B., A. Sjödin, M. Kurzon, S.A. Lederman, R.S. Jones, V. Rauh, L.L. Needham, D. Tang, M. Niedzwiecki, R.Y. Wang, and F. Perera. 2010. Prenatal exposure to PBDEs and neurodevelopment. Environ. Health Perspect. 118(5):712-719.

Ho, K.L., M.S. Yau, M.B. Murphy, Y. Wan, B.M. Fong, S. Tam, J.P. Giesy, K.S. Leung, and M.H. Lam. 2015. Urinary bromophenol glucuronide and sulfate conjugates: Potential human exposure molecular markers for polybrominated diphenyl ethers. Chemosphere 133:6-12.

Horn, S., and H. Heuer. 2010. Thyroid hormone action during brain development: More questions than answers. Mol. Cell Endocrinol. 315(1-2):19-26.

Ibhazehiebo, K., T. Iwasaki, J. Kimura-Kuroda, W. Miyazaki, N. Shimokawa, and N. Koibuchi. 2011. Disruption of thyroid hormone receptor-mediated transcription and thyroid hormone-induced purkinje cell dendrite arborization by polybrominated diphenyl ethers. Environ. Health Perspect. 119(2):168-175.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Review. Washington, DC: The National Academies Press.

Karpeta, A., J. Barc, A. Ptak, and E.L. Gregoraszczuk. 2013. The 2,2',4,4'-tetrabromodiphenyl ether hydroxylated metabolites 5-OH-BDE-47 and 6-OH-BDE-47 stimulate estradiol secretion in the ovary by activating aromatase expression. Toxicology 305:65-70.

Kilkenny, C., W.J. Browne, I.C. Cuthill, M. Emerson, and D.G. Altman. 2010. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biol. 8(6):e1000412.

Kim, K.H., D.D. Bose, A. Ghogha, J. Riehl, R. Zhang, C.D. Barnhart, P.J. Lein, and I.N. Pessah. 2011. Para- and ortho-substitutions are key determinants of polybrominated diphenyl ether activity toward ryanodine receptors and neurotoxicity. Environ. Health Perspect. 119(4):519-526.

Kodavanti, P.R., J.E. Royland, C. Osorio, W.M. Winnik, P. Oritz, L. Pei, R. Ramabhadran, and O. Alzate. 2015. Developmental exposure to a commercial PBDE mixture: Effects on protein networks in the cerebellum and hippocampus of rats. Environ. Health Perspect. 123(5):428-436.

Koenig, C.M., J. Lango, I.N. Pessah, and R.F. Berman. 2012. Maternal transfer of BDE-47 to offspring and neurobehavioral development in C57BL/6J mice. Neurotoxicol. Teratol. 34(6):571-580.

Konstantinova, A., G. Arsenault, B. Chittim, A. McAlees, R. McCrindle, D. Potter, C. Tashiro, and B. Yeo. 2008. Identification of the minor components of Great Lakes DE-71™ technical mix by means of 1H NMR and GD/MS. Chemosphere 73(1 Suppl.):S39-S43.

Lajeunesse, M.J. 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92(11):2049-2055.

Lam, J., P. Sutton, J. McPartland, L.I. Davidson, N. Daniels, S. Sen, D. Axelrad, B. Lanphear, D. Bellinger, and T.J. Woodruff. 2015. Applying the Navigation Guide Systematic Review Methodology, Case Study #5. Association Between Developmental Exposures to PBDEs and human Neurodevelopment: A Systematic Review of the Evidence Protocol, April 2015 [online]. Available: http://www.crd.york.ac.uk/PROSPEROFILES/17890_PROTOCOL_20150322.pdf [accessed August 3, 2016].

Lam, J., B. Lanphear, D. Bellinger, D.A. Axelrad, J. McPartland, P. Sutton, L. Davidson, N. Daniels, S. Sen, and T. Woodruff. 2016. Systematic Review and Meta-Analysis of Association between PBDE Exposure and IQ or ADHD. Draft Report, July 2016. University of California, San Francisco, CA.

Lam, J., B.P. Lanphear, D. Bellinger, D.A. Axelrad, J. McPartland, P. Sutton, L. Davidson, N. Daniels, S. Sen, and T.J. Woodruff. In press. Developmental PBDE exposure and IQ/ADHD in childhood: a systematic review and meta-analysis. Environ. Health Perspect.

Li, T., W. Wang, Y.W. Pan, L. Xu, and Z. Xia. 2013. A hudroxylated metabolite of flame-retardant PBDE-47 decreases the survival, proliferation and neuronal differentiation of primary cultured adult neural stem cells and interferes with signaling of ERK5 MAP kinase and neurotrophin 3. Toxicol. Sci. 134(1):111-124.

Lilienthal, H., A. Hack, A. Roth-Härer, S.W. Grande, and C.E. Talsness. 2006. Effects of developmental exposure to 2,2′,4,4′,5-pentabromodiphenyl ether (PBDE-99) on sex steroids, sexual development, and sexually dimorphic behavior in rats. Environ. Health Perspect. 114(2):194-201.

Lin, D.Y., H.R. Chao, Y.Y. Gou, and C.Y. Huang. 2010. Infants ingesting high breast milk levels of polybrominated diphenyl ethers may have negative impact on their neurodevelopment. Pp. 325-327 in 2010 International Conference on Chemistry and Chemical Engineering (ICCCE 2010) IEEE [online]. Available: http://ieeexplore.ieee.org/document/5560424/#full-text-section [accessed November 10, 2016].

Linares, V., M. Bellés, and J.L. Domingo. 2015. Human exposure to PBDE and critical evaluation of health hazards. Arch. Toxicol. 89(3):335-356.

Llansola, M., M. Hernandez-Viadel, S. Erceg, C. Montoliu, and V. Felipo. 2009. Increasing the function of the glutamate-nitric oxide-cyclic guanosine monophosphate pathway increases the ability to learn a Y-maze task. J. Neurosci. Res. 87(10):2351-2355.

Lorber, M. 2008. Exposure of Americans to polybrominated diphenyl ethers. J. Expo. Sci.Epidemiol. 18(1):2-19.

Meerts, I.A., J.J. Van Zanden, E.A. Luijks, I. Van Leeuwen-Bol, G. Marsh, E. Jakobsson, Å. Bergman, and A. Brouwer. 2000. Potent competitive interactions of some brominated flame retardants and related compounds with human transthyretin in vitro. Toxicol. Sci. 56(1):95-104.

Meerts, I.A., R.J. Letcher, S. Hoving, G. Marsh, A. Bergman, J.G. Lemmen, B. van der Burg, and A. Brouwer. 2001. In vitro estrogenicity of polybrominated diphenyl ethers, hydroxylated PBDEs, and polybrominated bisphenol A compounds. Environ. Health Perspect. 109(4):399-407.

Mercado-Feliciano, M., and R.M. Bigsby. 2008. Hydroxylated metabolites of the polybrominated diphenyl ether mixture DE-71 are weak estrogen receptor-alpha ligands. Environ. Health Perspect. 116(10):1315-1321.

Modesto, T., H. Tiemeier, R.P. Peeters, V.W. Jaddoe, A. Hofman, F.C. Verhulst, and A. Ghassabian. 2015. Maternal mild thyroid hormone insufficiency in early pregnancy and attention-deficit/hyperactivity disorder symptoms in children. JAMA Pediatr. 169(9):838-845.

Morgan, R.L., K.A. Thayer, L. Bero, N. Bruce, Y. Falck-Ytter, D. Ghersi, G. Guyatt, C. Hooijmans, M. Langendam, D. Mandrioli, R.A. Mustafa, E.A. Rehfuess, and A.A. Rooney. 2016. GRADE: Assessing the quality of evidence in environmental and occupational health. Environ. Int. 92-93:611-616.

Mundy, W. 2016. Presentation at Fortieth Annual Meeting of the Developmental Neurotoxicology Society, June 26, 2016, San Antonio, TX.

NCHS (National Center for Health Statistics). 2007. National Health and Nutrition Examination Survey 2003-2004 [online]. Available: https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/L28PBE_C.htm [accessed March 7, 2017].

NRC (National Research Council). 2014. Review of the Environmental Protection Agency's State-of-the Science Evaluation of Nonmontonic Dose-Response Relationships as They Apply to Endocrine Disruptors. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Pallocca, G., M. Grinberg, M. Henry, T. Frickey, J.G. Hengstler, T. Waldmann, A. Sachinidis, J. Rahnenführer, and M. Leist. 2016. Identification of transcriptome signatures and biomarkers specific for potential developmental toxicants inhibiting human neural crest cell migration. Arch. Toxicol. 90(1):159-180.

Ptak, A., G. Ludewig, H.J. Lehmler, A.K. Wojtowicz, L.W. Robertson, and E.L. Gregoraszczuk. 2005. Comparison of the actions of 4-chlorobiphenyl and its hydroxylated metabolites on estradiol secretion by ovarian follicles in primary cells in culture. Reprod. Toxicol. 20(1):57-64.

Ptak, A., G. Ludewig, M. Kapiszewska, Z. Magnowska, H.J. Lehmler, L.W. Robertson, E.L. Gregoraszczuk. 2006. Induction of cytochromes P450, caspase-3 and DNA damage by PCB3 and its hydroxylated metabolites in porcine ovary. Toxicol. Lett. 166(3):200-211.

Qiu, X., M. Mercado-Feliciano, R.M. Bigsby, and R.A. Hites. 2007. Measurement of polybrominated diphenyl ethers and metabolites in mouse plasma after exposure to a commercial pentabromodiphenyl ether mixture. Environ. Health Perspect. 115(7):1052-1058.

Reverte, I., A.B. Klein, J.L. Domingo, and M.T. Colomina. 2013. Long term effects of murine postnatal exposure to decabromodiphenyl ether (BDE-209) on learning and memory are dependent upon APOE polymorphism and age. Neurotoxicol. Teratol. 40:17-27.

Reverte, I., A. Pujol, J.L. Domingo, and M.T. Colomina. 2014. Thyroid hormones and fear learning but not anxiety are affected in adult apoE transgenic mice exposed postnatally to decabromodiphenyl ether (BDE-209). Physiol. Behav. 133:81-91.

Rice, D.C., W.D. Thompson, E.A. Reeve, K.D. Onos, M. Assadollahzadeh, and V.P. Markowski. 2009. Behavioral changes in aging but not young mice after neonatal exposure to the polybrominated flame retardant decaBDE. Environ. Health Perspect. 117(12):1903-1911.

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. Environ. Health Perspect. 122(7):711-718.

Rooney, A.A., G.S. Cooper, G.D. Jahnke, J. Lam, R.L. Morgan, A.L. Boyles, J.M. Ratcliffe, A.D. Kraft, H.J. Schunemann, and P. Schwingl. 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. Environ. Int. 92-93:617-629.

Roth, N., and M.F. Wilks. 2014. Neurodevelopmental and neurobehavioural effects of polybrominated and perfluorinated chemicals: A systematic review of the epidemiological literature using a quality assessment scheme. Toxicol. Lett. 230(2):271-281.

Roze, E., L. Meijer, A. Bakker, K.N. Van Braeckel, P.J. Sauer, and A.F. Bos. 2009. Prenatal exposure to organohalogens, including brominated flame retardants, influences motor, cognitive, and behavioral performance at school age. Environ. Health Perspect. 117(12):1953-1958.

Ryan, K.R., O. Sirenko, F. Parham, J.H. Hsieh, E.F. Cromwell, R.R. Tice, and M. Behl. 2016. Neurite outgrowth in human induced pluripotent stem cell-derived neurons as a high-throughput screen for developmental neurotoxicity or neurotoxicity. Neurotoxicology 53:271-281.

Sagiv, S.K., K. Kogut, F.W. Gaspar, R.B. Gunier, K.G. Harley, K. Parra, D. Villasenor, A. Bradman, N. Holland, and B. Eskenazi**.** 2015. Prenatal and childhood polybrominated diphenyl ether (PBDE) exposure and attention and executive function at 9-12 years of age. Neurotoxicol. Teratol. 52(Pt B):151-161.

Schmidt, B.Z., M. Lehmann, S. Gutbier, E. Nembo, S. Noel, L. Smirnova, A. Forsby, J. Hescheler, H.X. Avci, T. Hartung, M. Leist, J. Kobolák, and A. Dinnyés. 2017. In vitro acute and developmental neurotoxicity screening: an overview of cellular platforms and high-throughput technical possibilities. Arch. Toxicol. 91(1):1-33.

Siddiqi, M.A., R.H. Laessig, and K.D. Reed. 2003. Polybrominated diphenyl ethers (PBDEs): New pollutants-old diseases. Clin. Med. Res. 1(4):281-290.

Singh, S., A. Srivastava, V. Kumar, A. Pandey, D. Kumar, C.S. Rajpurohit, V.K. Khanna, S. Yadav, and A.B. Pant. 2016. Stem cells in neurotoxicology/developmental neurotoxicology: Current scenario and future prospects. Mol. Neurobiol. 53(10):6938-6949.

Stapleton, H.M., S.M. Kelly, R. Pei, R.J. Letcher, and C. Gunsch. 2009. Metabolism of polybrominated diphenyl ethers (PBDEs) by human hepatocytes in vitro. Environ Health Perspect. 117(2):197-202.

Staskal, D.F., H. Hakk, D. Bauer, J.J. Diliberto, and L.S. Birnbaum. 2006. Toxicokinetics of polybrominated diphenyl ether congeners 47, 99, 100, and 153 in mice. Toxicol. Sci. 94(1):28-37.

Sterne, J.A., M.A. Hernán, B.C. Reeves, J. Savović, N.D. Berkman, M. Viswanathan, D. Henry, D.G. Altman, M.T. Ansari, I. Boutron, J.R. Carpenter, A.W. Chan, R. Churchill, J.J. Deeks, A. Hróbjartsson, J. Kirkham, P. Jüni, Y.K. Loke, T.D. Pigott, C.R. Ramsay, D. Regidor, H.R. Rothstein, L. Sandhu, P.L. Santaguida, H.J. Schünemann, B. Shea, I. Shrier, P. Tugwell, L. Turner, J.C. Valentine, H. Waddington, E. Waters, G.A. Wells, P.F. Whiting, and J.P. Higgins. 2016. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ 355:i4919.

Ta, T.A., C.M. Koenig, M.S. Golub, I.N. Pessah, L. Qi, P.A. Aronov, and R.F. Berman. 2011. Bioaccumulation and behavioral effects of 2,2',4,4'-tetrabromodiphenyl ether (BDE-47) in perinatally exposed mice. Neurotoxicol. Teratol. 33(3):393-404.

Thuresson, K., P. Höglund, L. Hagmer, A. Sjödin, Å. Bergman, and K. Jakobsson. 2006. Apparent half-lives of hepta- to decabrominated diphenyl ethers in human serum as determined in occupationally exposed workers. Environ. Health Perspect.114(2):176-181.

Verma, P., P. Singh, and B.S. Gandhi. 2013. Prophylactic efficacy of *Bacopa monnieri* on decabromodiphenyl ether (PBDE-209)-induced alterations in oxidative status and spatial memory in mice. Asian J. Pharm. Clin. Res. 6(3):242-247.

Verma, P., P. Singh, and B.S. Gandhi. 2014. Neuromodulatory role of *Bacopa monnieri* on oxidative stress induced by postnatal exposure to decabromodiphenyl ether (PBDE -209) in neonate and young female mice. Iran. J. Basic Med. Sci. 17(4):307-311.

Viberg, H., A. Fredriksson, and P. Eriksson. 2003. Neonatal exposure to polybrominated diphenyl ether (PBDE 153) disrupts spontaneous behaviour, impairs learning and memory, and decreases hippocampal cholinergic receptors in adult mice. Toxicol. Appl. Pharmacol. 192(2):95-106.

Viberg, H., N. Johansson, A. Fredriksson, J. Eriksson, G. Marsh, and P. Eriksson. 2006. Neonatal exposure to higher brominated diphenyl ethers, hepta-, octa-, or nonabromodiphenyl ether, impairs spontaneous behavior and learning and memory functions of adult mice. Toxicol. Sci. 92(1):211-218.

Viberg, H., W. Mundy, and P. Eriksson. 2008. Neuonatal exposure to a decabrominated diphenyl ether (PBDE 209) results in changes in BDNF, CaMKII and GAP-43, biochemical substrates of neuronal survival, growth and synaptogenesis. Neurotoxicology 29(1):152-159.

Wang, X., L. Yang, Q. Wang, Y. Guo, N. Li, M. Ma, and B. Zhou. 2016. The neurotoxicity of DE-71: Effects on neural development and impairment of serotonergic signaling in zebrafish larvae. J. Appl. Toxicol. 36(12):1605-1613.

Wang, Z., L. Yang, Y. Wu, C. Huang, Q. Wang, J. Han, Y. Guo, Z. Shi, and B. Zhou. 2015. The developmental neurotoxicity of polybrominated diphenyl ethers: Effect of DE-71 on dopamine in zebrafish larvae. Environ. Toxicol. Chem. 34(5):1119-1126.

Westerink, R.H. 2014. Modulation of cell viability, oxidative stress, calcium homeostasis and votage- and ligand-gated ion channels as common mechanisms of action of (mixtures of) non-dioxin-like plychlorinated biphenyls and polybrominated diphenyl ethers. Environ. Sci. Pollut. Res. 21(10):6373-6383.

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

Wiseman, S.B., Y. Wan, H. Chang, X. Zhang, M. Hecker, P.D. Jones, and J.P. Giesy. 2011. Polybrominated diphenyl ethers and their hydroxylated/methoxylated analogs: Environmental sources, metabolic relationships, and relative toxicities. Mar. Pollut. Bull. 63(5-12):179-188.

Woodruff, T.J., and P. Sutton. 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(10):1007-1014.

Woods, R., R.O. Vallero, M.S. Golub, J.K. Suarez, T.A. Ta, D.H. Yasui, L.H. Chi, P.J. Kostyniak, I.N. Pessah, R.F. Berman, and J.M. LaSalle. 2012. Long-lived epigenetic interactions between perinatal PBDE exposure and Mecp2308 mutation. Hum. Mol. Genet. 21(11):2399-2411.

Zhang, H., X. Li, J. Nie, and Q. Niu. 2013. Lactation exposure to BDE-153 damages learning and memory, disrupts spontaneous behavior and induces hippocampus neuron death in adult rats. Brain Res. 1517:44-56.

Zhang, H., K. Yolton, G.M. Webster, A. Sjodin, A.M. Calafat, K.N. Dietrich, Y. Xu, C. Xie, J.M. Braun, B.P. Lanphear, and A. Chen. 2017. Prenatal PBDE and PCB exposures and reading, cognition, and externalizing behavior in children. Environ. Health Perspect. 125(4):746-752.

Zhao, W., J. Cheng, J. Gu, Y. Liu, M. Fujimura, and W. Wang. 2014. Assessment of neurotoxic effects and brain region distribution in rat offspring prenatally co-exposed to low doses of BDE-99 and methylmercury. Chemosphere 112:170-176.

5

Lessons Learned and Reflections on the Statement of Task

During the course of its deliberations, the committee developed a general strategy to collect and evaluate information regarding low-dose effects (see Chapter 2). The committee then applied two aspects of its strategy (investigation and analysis) to consider whether exposure to phthalates (see Chapter 3) or polybrominated diphenyl ethers (PBDEs; see Chapter 4) was associated with low-dose endocrine effects. The committee principally used systematic-review methods to complete and document these aspects. This decision was driven partly by the statement of task, which specified that the committee should complete systematic reviews of human and animal toxicology data for two or more chemicals that affect an endocrine hormone system. The committee's reliance on systematic-review methods is also consistent with recommendations made by previous National Academies committees (NRC 2014a,b).

The focus of the report on systematic review does not imply that it is the only evaluation tool needed to address all questions about low-dose toxicity. As described in Chapter 2, other options for investigation and analysis are available for use instead of or in conjunction with systematic review. Selection of the approaches will depend on the nature of the question and the potential health, social, and economic implications the answer will have. The committee strategy also describes potential options for future actions. For example, one potential action could be an update of an existing toxicity assessment. Although this type of action is consistent with the iterative strategy developed by the committee, completion of the additional steps needed to support this action were not pursued because the committee was not asked to complete risk assessments for any individual chemicals or classes of chemicals.

The purpose of this chapter is to provide additional discussion of lessons learned as they relate to committee efforts to address the statement of task and related issues concerning low-dose effects of endocrine active chemicals (EACs). It also provides lessons learned from performing the systematic reviews and integrating the human and animal evidence.

COMMITTEE STRATEGY

Development of a generic strategy for evaluating evidence of low-dose effects in Chapter 2 was deliberate and reflected the committee's desire to provide a framework that could be applied to many agents of concern regardless of the at-risk population, toxicity end point, or mechanism. Application of this generic strategy into one that meets the US Environmental Protection Agency's (EPA's) need to evaluate low-dose endocrine effects will require a combination of scientific and policy decisions on the agency's part. For example, implementation of an active surveillance program will likely necessitate that EPA identify specific EACs, dose ranges, populations, and end points to be monitored. The agency might need to perform multiple scoping exercises that lead to the development of *specific questions*. The problem formulation activities to address these questions will ultimately result in investigation and analyses tailored to address issues related to low-dose endocrine effects. Box 5-1 provides some examples of targeted analyses the committee performed to answer questions that arose during the course of working on the case example of phthalates in Chapter 3.

BOX 5-1 Examples of Targeted Analyses of Existing Data Performed by the Committee

During the course of performing the case examples presented in Chapters 3 and 4, several questions arose as the committee attempted to address the various elements of the statement of task. The approaches used by the committee to address these questions help illustrate the way the strategy proposed in Chapter 2 can be implemented. For illustration purposes, the committee has focused on questions related to phthalate effects on male reproductive-tract development. Each of these examples relies on a **targeted analysis of existing data** as the method of investigation; depending on the nature of the question and the potential action(s), additional methods of investigation and analysis might be required.

- *What is the relative sensitivity of humans to the effect of phthalates on testosterone production by the testes?* As discussed in Chapter 3, the human studies reviewed by the committee did not measure human testes testosterone production directly as was performed in the animal studies. The committee considered human-rodent xenograft data (a nontraditional end point) that could shed light on this question and conducted a literature review of relevant data and a meta-analysis of the subset of studies that evaluated testosterone production by human testes implanted in a rodent host (see Chapter 3 for additional details).
- *Are phthalate-induced effects on AGD adverse?* As discussed in Chapter 3, diethylhexyl phthalate (DEHP) is presumed to be a reproductive hazard to humans on the basis of effects on anogenital distance (AGD). The committee considered whether this effect could be considered adverse (see discussion later in this chapter). The committee performed a literature review of clinical literature evaluating the association of AGD with impaired reproductive performance in rodents and humans. Some of this literature evaluated adult populations or was independent of phthalate exposure and was therefore not part of the original research questions addressed by the committee's systematic review.
- *Are the human and animal dose-responses for phthalates similar?* As discussed in Chapter 3, it is difficult to compare the effects of different levels of phthalate exposure in animals and humans directly because animal studies typically report administered doses whereas studies in humans rely on the measurement of phthalate metabolites in urine or other body fluids. The committee considered human exposure and pharmacokinetic data that were found through literature reviews to address this question qualitatively. With respect to the committee strategy, answering this question more quantitatively might require the **generation of new data and models.**
- *Do phthalate effects on AGD or other reproductive end points demonstrate a non-monotonic dose response (NMDR)?* The committee's strategy can support investigations into whether certain chemicals and health effects might be associated with NMDRs. For example, the strategy is aligned with the recommendations of an earlier National Academies report that stressed the importance of performing systematic reviews of the literature to evaluate NMDRs (NRC 2014a). The committee evaluated the nature of the dose-response curve as part of its meta-analyses of animal and human studies on DEHP and AGD (see Chapter 3) and did not find evidence of an NMDR relationship. An important caveat to consider is that the studies evaluated might have lacked sufficient statistical power or a broad enough dose range to identify whether an NMDR exists.

CHARACTERIZING ADVERSITY

The committee discussed whether the effects of exposure to phthalate or PBDE exposure are adverse. The issue of adversity has been the subject of debate, and a number of definitions have been proposed (Kerlin et al. 2016). There have been recent attempts to develop clear criteria to evaluate whether an effect seen in nonclinical toxicology studies is adverse (Kerlin et al. 2016; Palazzi et al. 2016; Pandiri et al. 2017). The committee considered that guidance when drafting its definition of adverse.[1]

A determination of whether an effect is adverse requires expert judgment and should be based on evaluation of the effect in both the animal and human literature. For example, in the case of phthalate-induced effects, hypospadias was considered adverse because it represents a morphologic effect that

[1]Adverse effect: a biological change in an organism that results in an impairment of functional capacity, a decrease in the capacity to compensate for stress, or an increase in susceptibility to other influences (adapted from IPCS 2004).

might affect reproductive performance or behavior (Bubanj et al. 2004; Schlomer et al. 2014). Whether changes in either fetal testosterone concentrations or anogenital distance (AGD) are adverse has been the subject of discussion in the scientific community (Howdeshell et al. 2017). Several mechanisms, including androgen receptor antagonism and inhibition of androgen synthesis enzymes, can contribute to phthalate-induced reductions in fetal testosterone production (Howdeshell et al. 2017). In animal models of phthalate toxicity, multiple studies have found no effect on apical reproductive end points when treatment-related reduction in fetal testosterone was less than 40%, suggesting a point below which apical effects in studies in rodents are not observed (Gray et al. 2016). Likewise, a reduction in AGD is a biomarker of reduced in utero androgen concentrations (Thankamony et al. 2016) and might exhibit a similar threshold before changes in apical reproductive end points are seen. Several studies have reported that newborns born with hypospadias and cryptorchidism have shorter AGDs than infants without abnormalities (Hsieh et al. 2012; Jain and Singal 2013; Thankamony et al. 2014). In addition, several studies in adult males have reported that men who have reduced fertility—including lower sperm concentration, count, and motility—have shortened AGDs (Eisenberg et al. 2011, 2012; Mendiola et al. 2011, 2015; Eisenberg and Lipshultz 2015). At this time, the degree of AGD shortening necessary to observe apical end points remains unknown. It is therefore unclear that changes in testosterone or AGD in the absence of other apical reproductive end points would meet the first part of the committee's definition of an adverse effect—"a biological change in an organism that results in an impairment of functional capacity."

The committee also considered whether phthalate-induced changes in fetal testosterone or AGD might increase the susceptibility of an organism to other influences. Sufficient androgen activity is required for proper male reproductive-tract development, and this activity is a consequence of multiple inputs at the molecular level. In the rat model, data have shown that a phthalate-induced reduction in fetal testosterone made the animal more susceptible to the effects of another chemical (linuron) that also targeted the androgen signaling system, although the reduction by itself was insufficient to produce an adverse reproductive effect (Hotchkiss et al. 2004). On the basis of those data, a phthalate-related reduction in fetal testosterone or reduction in AGD would be considered an adverse effect because they meet another part of the definition: a biological change that results in "an increase in susceptibility to other influences."

The committee considered the PBDE-related effects on cognitive function to be adverse. The committee's systematic review supports the findings made by Lam et al. (in press) of an inverse association between developmental PBDE exposure and effects on IQ in children. Lam et al. reported a decrease of 3.7 IQ points per 10-fold increase in serum PBDE concentration. It is important to differentiate between effects seen at the individual and population levels (NRC 2009). First, the magnitude of response to PBDEs and other EACs can vary within a population because some individuals might be more affected and others less so by the same exposure (NRC 2009). At the individual level, large changes in IQ (e.g., 10 points) are considered adverse. Smaller changes on an individual level, such as a shift of 3-4 points, might not be associated with a functional impairment in cognitive function and are often within the range of variability when an individual is retested on an IQ test (Watkins and Smith 2013). However, given widespread exposure across the population, small changes in IQ could shift the entire population distribution in the direction of decreased function, which has quantifiable consequences integrated over the population, such as increased fraction of population with very low IQ or reduced aggregate economic output (Axelrad et al. 2007; Bellinger 2012). Therefore, PBDE-related reduction in IQ would be considered an adverse effect because they meet the first part of the definition: "a biological change in an organism that results in an impairment of functional capacity."

The committee's reasoning as to whether the changes produced by phthalate or PBDE exposure are adverse is likely to be applicable to other chemicals and end points. Some lessons learned from the committee's deliberation include the following:

- There is usually agreement regarding adversity of more severe outcomes because they fit into the first part of the definition concerning "impairment of functional capacity."

- Changes in continuous end points, such as hormone levels or biomarkers, might not lead to demonstrable "impairment of functional capacity" at the individual level. However, such effects might fit into the second or third parts of the definition of adversity relating to "a decrease in the capacity to compensate for stress" or "an increase in susceptibility to other influences." Decisions to label such changes as "adverse" would be strengthened by experimental data that demonstrate reduced compensatory capacity or increased susceptibility. In the absence of such data, scientific judgment (e.g., analogy to other end points) and policy considerations (e.g., the severity and magnitude of the possible effect) might be involved.
- The small magnitude of change of a continuous end point might not have demonstrable "impairment of functional capacity" at the individual level but could have quantifiable consequences on functional capacity when considered over the entire population. Such shifts in the population distribution might result in more individuals in the tail of the distribution or reduced aggregate function over the population. Therefore, when evaluating end points with respect to the first part of the definition, consideration needs to be given to effects over the population.

REFLECTIONS AND LESSONS LEARNED FROM THE SYSTEMATIC REVIEWS

Selection of Example Chemicals

The specification that the systematic reviews be performed on chemicals that "act through an endocrine-mediated pathway" and "that affect the estrogen, androgen, or perhaps other endocrine systems" led to extensive discussions about how stringent the committee should be about using mechanistic data to guide chemical selection. The committee's goal was to identify candidate chemicals that were presumed to be associated with endocrine effects. Although that approach seems straightforward, mechanisms are poorly understood for some of the candidate EACs that were initially considered. The committee eventually selected two chemical classes (phthalates and PBDEs) that had varying amounts of mechanistic information to illustrate different approaches that EPA might need to use when assessing low-dose endocrine effects. Other factors that influenced chemical selection were whether relevant animal and human data were available and the types of end points associated with each candidate chemical.

The committee also recognized that it might be advantageous for EPA to build on existing systematic reviews that are published in the peer-reviewed literature. During the course of its discussions, the committee became aware of a systematic review being conducted by Lam et al. (in press) on developmental exposures to PBDEs and human neurodevelopment. The willingness of Lam and coworkers to share an early draft of their study provided the committee with a unique opportunity for the committee to meet its objective of demonstrating how to build on a published systematic review. The committee remains indebted to Lam and coworkers for their generosity.

Methods

The committee initially performed a series of scoping exercises—conducting informal literature reviews and hosting a workshop—that helped define chemicals of interest, populations of concern, exposure windows, health end points, and other factors that led to the development of the specific research question and the appropriate PECO (Population, Exposure, Comparator, and Outcome) statements that guided the committee's investigation (Higgins and Green 2011; IOM 2011). Broadly stated, the phthalate research question asked whether in utero exposure to phthalates was associated with reproductive effects (as assessed by changes in fetal testosterone during gestation or at delivery, AGD, or the incidence of hypospadias) in male nonhuman mammals or humans. The PBDE research question asked whether developmental exposure to PBDEs was associated with neurobehavioral effects (as assessed by changes in learning, memory, attention, or response inhibition) in nonhuman mammals or humans. The systematic reviews performed by the committee were therefore hypothesis driven and designed to answer a set of focused questions. Although the project was intended to address issues surrounding low-dose effects, the

committee did not constrain exposure to low dose in the PECO statement and therefore did not use it as an eligibility criterion. That decision was a deliberate choice because the committee did not want to a priori constrain the investigation by any preconceived notion of low dose. Instead, the committee chose to address whether effects occurred at a low dose as a separate subsequent step in the process.

The committee discussed at length whether it could provide EPA with advice about when a systematic review should be performed but decided it could not be more specific because that decision will depend on the availability of data and resources, the anticipated actions, the time frame for decision making, and other factors. As the committee can attest, one disadvantage in conducting a systematic review is that it can be time and resource intensive, particularly for individuals that have not previously conducted a systematic review. Some steps are inherently resource consuming. For example, two individuals independently perform most steps in the systematic-review process, such as data abstraction and risk of bias determinations. The committee therefore recognized that there was a need for the inclusion of investigation methods that do not rely on systematic-review methods. One form of accelerated evidence synthesis that has been suggested by practitioners of systematic review is a rapid review, in which components of the systematic-review process are simplified or omitted (e.g., the need for two independent reviewers) to produce information in a more timely manner (Khangura et al. 2014; Polisena et al. 2015; Tricco et al. 2015). The committee recognizes the need to streamline the process; however, evidence suggests that rapid reviews should not be viewed as a substitute for a systematic review and that the time savings correlate with decreased methodologic quality or robustness (Harker and Kleijnen 2012; Featherstone et al. 2015). For some questions, systematic review methods are not expected to be time intensive (e.g., questions with small literature bases or data-poor chemicals). For other questions, the ability to streamline, increase efficiency, or automate portions of the review process might be considered by EPA. Efforts are currently under way to automate individual tasks in a systematic review, such as the literature search, study selection, and data extraction (Tsafnat et al. 2014; Jonnalagadda et al. 2015). The committee recognizes that the methods and role of systematic review and meta-analysis in toxicology are evolving rapidly and EPA will need to stay abreast of these developments, strive for transparency, and use appropriate methods to address its questions.

The committee used the National Toxicology Program's Office of Health Assessment and Translation (OHAT) method (Rooney et al. 2014; NTP 2015) to evaluate the confidence in the evidence. Overall, the committee found the OHAT method was relatively easy to implement but identified a few challenges. For example, a body of evidence from experimental animal studies was given an initial confidence rating of "high" on the basis of the study-design issues that are inherent in experimental studies. That approach is consistent with the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach for considering randomized controlled trials. Multiple factors were then considered to determine whether to downgrade (e.g., risk of bias concern or unexplained inconsistency in the results) or upgrade (e.g., large magnitude of effect or evidence of a dose-response relationship) confidence in the body of evidence. In practice, the committee found that bodies of evidence had multiple risk of bias issues and found it challenging to determine whether to downgrade confidence by one or more levels on the basis of the problems with study design or conduct observed. The method would benefit from clearer guidelines or examples to illustrate the consideration of factors for downgrading or upgrading confidence.

Several other lessons became apparent during the course of the study. They include the need for improved reporting of study design, conduct, and results on the part of scientists to facilitate risk of bias evaluations; approaches for displaying behavioral and categorical data within a systematic review; and inclusion of subject-matter expertise on the review team, including expertise in meta-analysis and other statistical approaches. It is often not clear if a study used best practices in the study design and conduct but failed to report those good practices, or if there were issues in study conduct that would result in potential bias. Researchers should be encouraged to follow reporting guidelines (e.g., the ARRIVE guidelines for reporting of animal studies and the STROBE statement for reporting of observational studies in epidemiology). They should also use methods to minimize bias in research conduct particularly for issues where there is empirical evidence that risk of bias practices can affect the effect size (e.g., the key issues of randomizing animals to treatment or and blinding of outcome assessors to study groups).

The committee also found that quantitative data analysis methods can be valuable in evaluating a body of evidence. For phthalates, the use of common end-point measures across most studies enabled extensive use of meta-analysis techniques, the results of which contributed to evaluations of the precision, heterogeneity, magnitude of effect, and presence of a dose-response gradient across studies. Use of meta-analysis techniques is much more robust than relying on individual study results, because they account for potential heterogeneity across studies and the statistical power of each individual study. For PBDEs, the variety of test methods and end points limited the use meta-analysis, although the committee did demonstrate its use for one of the more commonly reported end points. In this case, the committee's preferred option would have been to conduct benchmark dose modeling on individual studies; however, many studies did not report sufficient details on the study design (e.g., group sizes) or results (e.g., standard deviations) to estimate benchmark doses. Thus, the committee had to rely on reported lowest-observed-adverse-effect levels (LOAELs) and no-observed-adverse-effect levels (NOAELs) based on pair-wise statistical significance in evaluating studies. The use of LOAELs and NOAELs is less than ideal because they depend highly on individual study-design characteristics; therefore, apparent differences among studies might be explained by design differences, such as sample size or dose spacing, rather than true inconsistency. Indeed, the committee found that some of its concerns about consistency and heterogeneity in the PBDE studies were ameliorated by the results of the meta-analysis performed on the Morris water maze data.

As discussed above and elsewhere in the report, the committee used existing methods and software for performing the systematic reviews, the meta-analyses, and the evidence integration. The selected methods and software were primarily chosen because of the committee's experience with using them and should not be taken as an indication of an endorsement of them or that they are the recommended approaches for EPA to use. The committee recognizes that other software, statistical packages, and evidence integration approaches could fit EPA's needs.

LESSONS LEARNED FROM EVIDENCE INTEGRATION

In keeping with the statement of task, the committee demonstrated how human and animal data streams can be integrated and used to determine whether a likely causal association is supported by the evidence. Determinations regarding causality were constrained by the PECO statements that were described earlier. The committee synthesized the animal and human evidence and reached hazard conclusions using an OHAT framework (NTP 2015). Earlier chapters describe the data integration steps in more detail. Using the OHAT framework and language used to describe the strength of the evidence, the committee reached a number of conclusions regarding the endocrine toxicity of phthalates and PBDEs. Two example causality statements are provided below.

- Diethylhexyl phthalate (DEHP) is presumed to be a reproductive hazard to humans, and there is moderate evidence that decreased AGD in humans occurs following low-dose exposure to this phthalate.
- Developmental exposure to BDE-47 is presumed to pose a hazard to intelligence in humans, and there is a moderate level of evidence that effects on IQ occur following low-dose exposure to this congener.

The Use of Mechanistic Data for Evidence Integration

The committee considered mechanistic data, which included pharmacokinetic information, in reaching its final hazard identification conclusions for each end point. For example, the committee used mechanistic evidence to support the finding that DEHP effects on AGD in humans were biologically plausible. The OHAT hazard identification scheme also allows consideration of mechanistic data to upgrade or downgrade the initial hazard determination. The committee found that the guidance in the OHAT handbook on the level of evidence needed to upgrade or downgrade the initial hazard determination on the ba-

sis of mechanistic data was somewhat lacking, and there are few OHAT monographs or published systematic reviews that have used the approach. The method would benefit from additional clarity in the guidelines or particularly from examples to illustrate mechanistic evidence that would provide strong support for biological plausibility of the observed effect sufficient to justify upgrading or downgrading the hazard conclusion.

Mechanistic data were also used to evaluate whether there was dose-response concordance between humans and animals. In the case of DEHP effects on AGD, the committee noted significant species differences in phthalate metabolism and clearance, and effects on fetal testosterone (Ito et al 2005; Gaido et al. 2007; Johnson et al. 2012). In addition, qualitative and quantitative differences in phthalate pharmacokinetics also occur between rodents, nonhuman primates, and humans (Kessler et al. 2004; McKinnell et al. 2009; Kurata et al. 2012). The committee found comparing evidence on dose-response relationships between animal and human studies to be challenging and imprecise because animal studies often measure external administered doses (usually without measures of internal dose), whereas human studies measure biomarkers of internal dose (with estimates of the external administered dose being uncertain). Toxicology studies that measure internal dose metrics, including metrics that are similar to those used in human biomonitoring and those most relevant to the target tissue dose, could help address those challenges.

Broader mechanistic questions concerning how EACs might alter normal hormone function at low doses have been raised in the scientific literature (Skakkebaek et al. 2011; Vandenberg 2014; Maqbool et al. 2016). Although those questions are potentially important for risk assessment of EACs, they were deemed beyond the charge of the committee.

End-Point Concordance

The methods used to assess fetal testosterone, AGD, and hypospadias were qualitatively similar between animal and human studies, and this simplified the analysis of whether the responses seen were concordant or discordant. In addition, end-point consistency between human and animal studies provided additional confidence in biological plausibility for hazard identification. Moreover, because similar end points were evaluated in multiple studies, meta-analyses were possible for several end points of interest.

In the case of PBDE effects on cognitive function, change in IQ was a primary measure in children, and a wide array of neurobehavioral assays were used in the animal studies. Although the assays evaluate cognitive function in animals (e.g., tests of learning and memory), the types of tests used, the timing of evaluation during postnatal development, and other methodologic differences restricted the ability of the committee to synthesize the data using meta-analysis or other methods.

OTHER ISSUES

A related element of the statement of task asked the committee to consider adverse outcome pathways (AOPs) and high-throughput data in its analysis. The committee found that using AOPs and high-throughput data was difficult because molecular initiating events involved in the phthalate reproductive effects remain unclear, and steroidogenic assays that are used in current high-throughput assay systems (e.g., ToxCast) often rely on human adrenal cell lines of unknown mechanistic relevance for phthalate toxicity. In the case of the PBDEs, several potential mechanisms have been proposed (Costa et al. 2014); however, none have been conclusively linked to the neurobehavioral outcomes evaluated by the committee. Although high-throughput data were not helpful in the committee's analyses of phthalates and PBDE, such data could be used for priority setting and other uses. For example, chemicals that exhibit endocrine activity in a high-throughput assay might be given a higher priority for future testing. That approach is consistent with the EPA's Endocrine Disruptor Screening Program. High-throughput data might be used to differentiate chemicals on the basis of bioactivity potency to set priorities for chemical testing. High-throughput data might also be used to support read-across methods for chemicals that have few human or animal data. In addition, the use of reverse toxicokinetic methods to support in vitro–to–in vivo extrapolations that can be used to convert an in vitro concentration into an estimated serum concentration will be

important (Wetmore et al. 2012). The committee anticipates that methods will emerge to support the use of in vitro data streams in systematic reviews.

The committee also considered whether animal toxicity studies could predict the low-dose effects seen with phthalates or PBDEs in people. The committee found that, although animal data could help identify hazards associated with phthalates and PBDE exposure, they were unable to predict exposures at which effects occurred in people. Indeed, differences in exposure between animal studies and those observed in the general human population spanned several orders of magnitude. Strategies to help resolve apparent discrepancies often rest on an improved understanding of health effects seen in people, revisions to animal-testing guidelines that help improve their predictive value, consideration of animal studies that include additional end points, and efforts to use mechanistic and pharmacokinetic data to bridge these seemingly disparate data streams (NRC 2007, 2009). The recent National Academies report *Using 21st Century Science to Improve Risk-Related Evaluations* (NASEM 2017) emphasized the need to align environmental and test-system exposures and develop the models and methods necessary to do so. Thus, new test methods, models, and approaches will need to evolve to address the apparent discrepancies.

REFERENCES

Axelrad, D.A., D.C. Bellinger, L.M. Ryan, and T.J. Woodruff. 2007. Dose-response relationship of prenatal mercury exposure and IQ: An integrative analysis of epidemiologic data. Environ. Health Perspect. 115(4):609-615.

Bellinger, D.C. 2012. A strategy for comparing the contributions of environmental chemicals and other risk factors to neurodevelopment of children. Environ. Health Perspect. 120(4):501-507.

Bubanj, T.B., S.V. Perovic, R.M. Milicevic, S.B. Jovcic, Z.O. Marjanovic, and M.M. Djordjevic. 2004. Sexual behavior and sexual function of adults after hypospadias surgery: A comparative study. J. Urol. 171(5):1876-1879.

Costa, L.G., R. de Laat, S. Tagliaferri, and C. Pellacani. 2014. A mechanistic view of polybrominated diphenyl ether (PBDE) developmental neurotoxicity. Toxicol. Lett. 230(2):282-294.

Eisenberg, M.L., and L.I. Lipshultz. 2015. Anogenital distance as a measure of human male fertility. J. Assist. Reprod. Genet. 32(3):479-484.

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(5):0018973.

Eisenberg, M.L., M. Shy, R.C. Walters, and L.I. Lipshultz. 2012. The relationship between anogenital distance and azoospermia in adult men. Int. J. Androl. 35(5):726-730.

EPA (U.S. Environmental Protection Agency). 2013. State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures. U.S. Environmental Protection Agency. June 2013 [online]. Available: https://www.epa.gov/sites/production/files/2016-01/documents/nmdr.pdf [accessed November 11, 2015].

Featherstone, R.M., D.M. Dryden, M. Foisy, J.M. Guise, M.D. Mitchell, R.A. Paynter, K.A. Robinson, C.A. Umscheid, and L. Hartling. 2015. Advancing knowledge of rapid reviews: An analysis of results, conclusions and recommendations from published review articles examining rapid reviews. Syst. Rev. 17(4):50.

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. Toxicol. Sci. 97(2):491-503.

Gray, L.E., Jr., J. Furr, K.R. Tatum-Gibbs, C. Lambright, H. Sampson, B.R. Hannas, V.S. Wilson, A. Hotchkiss, and P.M. Foster. 2016. Establishing the "biological relevance" of dipentyl phthalate reductions in fetal rat testosterone production and plasma and testis testosterone levels. Toxicol. Sci. 149(1):178-191.

Harker, J., and J. Kleijnen. 2012. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. Int. J. Evidence Based Healthcare 10(4):397-410.

Higgins, J.P.T., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration [online]. Available: www.handbook.cochrane.org [accessed June 1, 2017].

Hotchkiss, A.K., L.G. Parks-Saldutti, J.S. Ostby, C. Lambright, J. Furr, J.G. Vandenbergh, and L.E. Gray, Jr. 2004. A mixture of the "antiandrogens" linuron and butyl benzyl phthalate alters sexual differentiation of the male rat in a cumulative fashion. Biol. Reprod. 71(6):1852-1861.

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. Int. J. Hyg. Environ. Health 220(2 Pt. A):179-188.

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. Hum. Reprod. 27(6):1577-1580.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Review. Washington, DC: The National Academies Press.

IPCS (International Programme on Chemical Safety). 2004. IPCS Risk Assessment Terminology. Part 1: IPCS/OECD Key Generic Terms Used in Chemical Hazard/Risk Assessment. Part 2: IPCS Glossary of Key Exposure Assessment Terminology. Geneva: World Health Organization [online]. Available: http://www.inchem.org/documents/harmproj/harmproj/harmproj1.pdf [accessed November 11, 2015].

Ito, Y., H. Yokota, R. Wang, O. Yamanoshita, G. Ichihara, H. Wang, Y. Kurata, K. Takagi, and T. Nakajima. 2005. Species differences in the metabolism of di(2-ethylhexyl) phthalate (DEHP) in several organs of mice, rats, and marmosets. Arch. Toxicol. 79(3):147-154.

Jain, V.G., and A.K. Singal. 2013. Shorter anogenital distance correlates with undescended testis: A detailed genital anthropometric analysis in human newborns. Hum. Reprod. 28(9):2343-2349.

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. Toxicol. Sci. 129(2):235-248.

Jonnalagadda, S.R., P. Goyal, and M.D. Huffman. 2015. Automating data extraction in systematic reviews: A systematic review. Syst. Rev. 4:78.

Kerlin, R., B. Bolon, J. Burkhardt, S. Francke, P. Greaves, V. Meador, and J. Popp. 2016. Scientific and Regulatory Policy Committee: Recommended ("best") practices for determining, communicating, and using adverse effect data from nonclinical studies. Toxicol. Pathol. 44(2):147-162.

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. Toxicol. Appl. Pharmacol. 195(2):142-153.

Khangura, S., J. Polisena, T.J. Clifford, K. Farrah, and C. Kamel. 2014. Rapid review: An emerging approach to evidence synthesis in health technology assessment. Int. J. Technol. Assess. Health Care 30(1):20-27.

Kurata, Y., F. Makinodan, N. Shimamura, and M. Katoh. 2012. Metabolism of di (2-ethylhexyl) phthalate (DEHP): Comparative study in juvenile and fetal marmosets and rats. J. Toxicol. Sci. 37(1):33-49.

Lagarde, F., C. Beausoleil, S.M. Belcher, L.P. Belzunces, C. Emond, M. Guerbet, and C. Rousselle. 2015. Non-monotonic dose-response relationships and endocrine disruptors: A qualitative method of assessment. Environ. Health 14(1):13.

Lam, J., B.P. Lanphear, D. Bellinger, D.A. Axelrad, J. McPartland, P. Sutton, L. Davidson, N. Daniels, S. Sen, and T.J. Woodruff. In press. Developmental PBDE exposure and IQ/ADHD in childhood: a systematic review and meta-analysis. Environ. Health Perspect.

Maqbool, F., S. Mostafalou, H. Bahadar, and M. Abdollahi. 2016. Review of endocrine disorders associated with environmental toxicants and possible involved mechanisms. Life Sci. 145:265-273.

McKinnell, C., R.T. Mitchell, M. Walker, K. Morris, C.J.H. Kelnar, W.H. Wallace, and R.M. Sharpe. 2009. Effect of fetal or neonatal exposure to monobutyl phthalate (MBP) on testicular development and function in the marmoset. Hum. Reprod. 24(9):2244-2254.

Mendiola, J., R.W. Stahlhut, N. Jorgensen, F. Liu, and S.H. Swan. 2011. Shorter anogenital distance predicts poorer semen quality in young men in Rochester, New York. Environ. Health Perspect. 119(7):958-963.

Mendiola, J., M. Melgarejo, M. Monino-Garcia, A. Cutillas-Tolin, J.A. Noguera-Velasco, and A.M. Torres-Cantero. 2015. Is anogenital distance associated with semen quality in male partners of subfertile couples? Andrology 3(4):672-676.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2017. Using 21st Century Science to Improve Risk-Related Evaluations. Washington, DC: The National Academies Press.

NRC (National Research Council). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

NRC. 2014a. Review of the Environmental Protection Agency's State-of-the Science Evaluation of Nonmontonic Dose-Response Relationships as They Apply to Endocrine Disruptors. Washington, DC: The National Academies Press.

NRC. 2014b. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015. Handbook for Conducting a Literature-based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Palazzi, X., J.E. Burkhardt, H. Caplain, V. Dellarco, P. Fant, J.R. Foster, S. Francke, P. Germann, S. Gröters, T. Harada, J. Harleman, K. Inui, W. Kaufmann, B. Lenz, H. Nagai, G. Pohlmeyer-Esch, A. Schulte, M. Skydsgaard, L. Tomlinson, C.E. Wood, and M. Yoshida. 2016. Characterizing "adversity" of pathology findings in nonclinical toxicity studies: Results from the 4th ESTP International Expert Workshop. Toxicol. Pathol. 44(6):810-824.

Pandiri, A.R., R.L. Kerlin, P.C. Mann, N.E. Everds, A.K. Sharma, L.P. Myers, and T.J. Steinbach. 2017. Is it adverse, nonadverse, adaptive, or artifact? Toxicol. Pathol. 45(1):238-247.

Polisena, J., C. Garritty, C.A. Umscheid, C. Kamel, K. Samra, J. Smith, and A. Vosilla. 2015. Rapid Review Summit: An overview and initiation of a research agenda. Syst. Rev. 4:137.

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 science assessments. Environ. Health Perspect. 122(7):711-718.

Schlomer, B., B. Breyer, H. Copp, L. Baskin, and M. DiSandro. 2014. Do adult men with untreated hypospadias have adverse outcomes? A pilot study using a social media advertised survey. J. Pediatr. Urol. 10(4):672-679.

Skakkebaek, N.E., J. Toppari, O. Söder, C.M. Gordon, S. Divall, and M. Draznin. 2011. The exposure of fetuses and children to endocrine disrupting chemicals: A European Society for Paediatric Endocrinology (ESPE) and Pediatric Endocrine Society (PES) call to action statement. J. Clin. Endocrinol. Metab. 96(10):3056-3058.

Thankamony, A., N. Lek, D. Carroll, M. Williams, D.B. Dunger, C.L. Acerini, K.K. Ong, and I.A. Hughes. 2014. Anogenital distance and penile length in infants with hypospadias or cryptorchidism: Comparison with normative data. Environ. Health Perspect. 122(2):207-211.

Thankamony, A., V. Pasterski, K.K. Ong, C.L. Acerini, and I.A. Hughes. 2016. Anogenital distance as a marker of androgen exposure in humans. Andrology 4(4):616-625.

Tricco, A.C., J. Antony, W. Zarin, L. Strifler, M. Ghassemi, J. Ivory, L. Perrier, B. Hutton, D. Moher, and S.E. Straus. 2015. A scoping review of rapid review methods. BMC Med. 13:224.

Tsafnat, G., P. Glasziou, M.K. Choong, A. Dunn, F. Galgani, and E. Coiera. 2014. Systematic review automation technologies. Syst. Rev. 3:74.

Vandenberg, L.N. 2014. Low-dose effects of hormones and endocrine disruptors. Vitam. Horm. 94:129-165.

Watkins, M.W., and L.G. Smith. 2013. Long-term stability of the Wechsler Intelligence Scale for Children–Fourth Edition. Psychol. Assess. 25(2):477-483.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

Appendix A

Biosketches of the Committee on Endocrine-Related Low-Dose Toxicity

David C. Dorman (*Chair*) is a professor of toxicology in the Department of Molecular Biomedical Sciences at North Carolina State University. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has chaired or served on several National Research Council (NRC) committees and is a National Associate of the NRC. He has served on other advisory boards for the US Navy, the National Aeronautics and Space Administration, the US Department of Agriculture, and is a former member of the National Toxicology Program (NTP) Board of Scientific Counselors. Dr. Dorman is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Sciences. He received a DVM from Colorado State University. He completed a combined PhD and veterinary toxicology residency program at the University of Illinois at Urbana-Champaign, and is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.

Weihsueh Chiu is a professor in the Department of Veterinary Integrative Biosciences in the College of Veterinary Medicine and Biomedical Sciences at Texas A&M University. Before joining the university, he worked at the US Environmental Protection Agency (EPA) for more than 14 years, most recently as chief of the Toxicity Pathways Branch in the Integrated Risk Information System (IRIS) Division of the National Center for Environmental Assessment. His research has focused on human health risk assessment, particularly with respect to toxicokinetics, mechanisms of toxicity, physiologically based pharmacokinetic modeling, dose-response assessment, and characterizing uncertainty and variability. Dr. Chiu led the development of the EPA's 2011 IRIS assessment of trichloroethylene, which pioneered the use of probabilistic methods for characterizing uncertainty and variability in toxicokinetics and dose response. He was a member of the NRC's Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures. Dr. Chiu received a PhD in physics from Princeton University.

Barbara F. Hales is a James McGill Professor in the Department of Pharmacology and Therapeutics at McGill University. Her research interests are in the mechanisms of action of drugs as teratogens. She studies developmental toxicity using a combination of in vivo, in vitro, and molecular approaches with the goal of elucidating how the embryo responds to insult after direct or maternal exposure and the consequences to progeny of paternal drug exposure. Dr. Hales is a past president of the Teratology Society and is currently co-chair of the Chemicals Management Plan Science Committee of the Government of Canada. She received an MSc in pharmacognosy from the Philadelphia College of Pharmacy and Science and a PhD in pharmacology and therapeutics from McGill University.

Russ B. Hauser is the Frederick Lee Hisaw Professor of Reproductive Physiology and professor of environmental and occupational epidemiology in the Department of Environmental Health at the Harvard T.H. Chan School of Public Health. He also holds an appointment at the Harvard Medical School, where he is professor of obstetrics, gynecology, and reproductive biology. Dr. Hauser's research focuses on the health risks posed by exposure to environmental chemicals that adversely affect human development and reproductive health. He has served on several NRC and Institute of Medicine committees, including the

Committee to Review EPA's State of the Science Paper on Nonmonotonic Dose Response and the Committee on the Health Risks of Phthalates. Dr. Hauser is a member of two EPA Science Advisory Boards. He served on the US Consumer Product Safety Commission's Chronic Hazard Advisory Panel examining the effects of phthalates on children's health. He received an MD from the Albert Einstein College of Medicine and an MPH and a ScD from the Harvard School of Public Health.

Kamin J. Johnson is a lead scientist for The Dow Chemical Company's Toxicology and Environmental Research and Consulting function, responsible for the scientific conduct and interpretation of developmental and reproductive toxicology studies. He has served on study sections of the National Institutes of Health reviewing reproductive toxicology grants, and he was a counselor for the Reproductive and Toxicology Specialty Section of the Society of Toxicology. His research interests are in the molecular and cellular biology of fetal and postnatal testis function, as well as mechanisms of testicular toxicants. Dr. Johnson received a PhD in molecular biology, cell biology, and biochemistry from Brown University.

Karen A. Robinson is an associate professor at the Johns Hopkins University School of Medicine. She also serves as director of the Johns Hopkins University Evidence-based Practice Center and is a member of the core faculty in the Center for Clinical Trials and Evidence Synthesis at the university's Bloomberg School of Public Health. Dr. Robinson's research focuses on evidence-based health care and evidence-based research. She conducts systematic reviews that are used to develop clinical practice guideline and to inform other health decisions. Dr. Robinson received an MSc in health sciences from the University of Waterloo, Ontario, and a PhD in epidemiology from the Johns Hopkins Bloomberg School of Public Health.

Andrew A. Rooney is deputy director of the Office of Health Assessment and Translation (OHAT) in the National Toxicology Program at the National Institute of Environmental Health Sciences (NIEHS). He has been developing risk assessment methods and guidance throughout his professional career and is a principal author of the 2012 World Health Organization/International Programme on Chemical Safety Guidance for Immunotoxicity Risk Assessment for Chemicals. Most recently, Dr. Rooney has been working on emerging issues in toxicology and environmental health, including methods to address study quality in terms of risk of bias for human, animal, and mechanistic studies and adaptation of systematic review methods for addressing environmental health questions. He led the team that developed the OHAT approach to systematic review. Dr. Rooney has an MS and a PhD in zoology from the University of Florida.

Ruthann Rudel is director of research at the Silent Spring Institute, an independent not-for-profit environmental research organization focused on women's health. She leads the exposure and toxicology research program focusing on endocrine active chemicals and on mechanisms by which chemicals may influence breast cancer. Ms. Rudel's work in toxicology includes a review of early life exposure to chemicals that alter mammary gland development and implications for testing protocols and risk assessment. She also has an appointment as a research associate in the Department of Pathology and Laboratory Medicine at Brown University. Ms. Rudel was a member of the National Toxicology Program's Board of Scientific Counselors and of the Society of Toxicology's Regulatory Affairs and Legislative Assistance Committee. She received an MS in environmental management and policy from Tufts University.

Sheela Sathyanarayana is an associate professor in the Department of Pediatrics and an adjunct associate professor in the Department of Environmental and Occupational Health Sciences at the University of Washington. She is also an attending physician at Harborview Medical Center and Seattle Children's Hospital. Her research interests focus on exposures to endocrine disrupting chemicals, including phthalates and bisphenol A, and their effects on reproductive development. Currently, Dr. Sathyanarayana is the center director and clinical director for The Infant Development and Environment Study, which is a multicenter cohort study of phthalate exposures in pregnancy and health outcomes in children. She is a former chair of the EPA's Children's Health Protection Advisory Committee. Dr. Sathyanarayana earned

an MD from the University of Southern California and an MPH in epidemiology from the University of Washington.

Susan L. Schantz is a professor of toxicology in the Department of Comparative Biosciences, College of Veterinary Medicine, at the University of Illinois at Urbana-Champaign. She is also director of a NIEHS T32 training program in endocrine, developmental, and reproductive toxicology and director of a Children's Environmental Health Research Center jointly funded by the NIEHS and the EPA. In addition, she is currently the interim director of the Neuroscience Program. Dr. Schantz's research interests involve understanding the neurobehavioral effects of chemical exposures during development and aging. She conducts research in both laboratory-based animal studies and parallel epidemiologic studies. She has served as president of the Neurotoxicology Specialty Section of the Society of Toxicology and president of the Neurobehavioral Teratology Society. Dr. Schantz was also a member of the NRC's Committee to Assess the Health Implications of Perchlorate Ingestion. She received a PhD in environmental toxicology from the University of Wisconsin–Madison.

Katrina Waters is deputy director of the Biological Sciences Division at the Pacific Northwest National Laboratory. Her research interests are focused on the integration of genomics, proteomics, metabolomics, and high-throughput screening data to enable predictive mechanistic modeling of disease and toxicity pathways. She has served on the EPA's Board of Scientific Counselors Subcommittee on Chemical Safety for Sustainability and the US Food and Drug Administration's Scientific Advisory Board to the National Center for Toxicological Research. She served on the NRC's Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures. Dr. Waters received a PhD in biochemistry from the University of Wisconsin–Madison, and did a postdoctoral fellowship on endocrine disruptors at the Chemical Industry Institute of Toxicology.

Appendix B

Workshop on Potential Case Studies for Unraveling Endocrine-Related Low-Dose Toxicity

National Academies of Sciences, Engineering, and Medicine
500 Fifth Street NW, Room 100
Washington, DC

February 3, 2016

8:00 **Registration**

8:30 **Welcome and Goals of the Workshop**
David Dorman, Committee Chair

8:45 **Case Example 1: Phthalates and Male Reproductive Malformations**
Moderators: Kamin Johnson, Russ Hauser, Sheela Sathyanarayana
Panelists:
- *Kim Boekelheide, Brown University*
- *Jodi Flaws, University of Illinois at Urbana-Champaign (via teleconference)*
- *Earl Gray, U.S. Environmental Protection Agency*
- *Bernard Jégou, National Institute of Health and Medical Research (France) (via teleconference)**
- *John Meeker, University of Michigan (via teleconference)*

10:30 **Break**

10:45 **Case Example 2: TCDD and Male Reproductive Effects**
Moderators: Russ Hauser, Kamin Johnson, Andrew Rooney
Panelists:
- *Michael DeVito, National Institute of Environmental Health Sciences**
- *Brenda Eskenazi, University of California, Berkeley (via teleconference)*
- *Earl Gray, U.S. Environmental Protection Agency*

12:30 **Break (Cafeteria on Third Floor)**

1:30 **Case Example 3: Bisphenol A and Female Reproductive Effects**
Moderators: Weihsueh Chiu, Katrina Waters, Karen Robinson
Panelists:
- *Joseph Braun, Brown University*
- *Daniel Doerge, U.S. Food and Drug Administration*
- *Jodi Flaws, University of Illinois at Urbana-Champaign (via teleconference)*

3:15 **Break**

3:30 **Open Microphone**
Each speaker has a maximum time limit of 5 minutes. Accompanying written materials are encouraged.

4:00 **Adjourn**

*This individual was unable to participate on the day of the workshop.

WORKSHOP HANDOUTS

PECO Statements for Phthalates

(1) Human Study Question: Is in utero exposure to phthalates in humans associated with male reproductive malformations?

TABLE 1 Human PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Males without restriction based on age
Exposure	In utero exposure to one or more of the following phthalates: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), and/or dipentyl phthalate (CAS no. 131-18-0). When exposure data to more than one of the selected phthalates are obtained, the exposure data will be considered as a cumulative exposure using appropriate potency factors for individual phthalate congeners. No restrictions based on route of exposure, based on biomonitoring data (e.g., urine, blood, or other specimens), environmental measures (e.g., air or water concentrations), or indirect measures (e.g., job title).
Comparators	Populations exposed at lower levels of the selected phthalates
Outcomes	**Primary outcomes:** Male reproductive effects, including alterations in fertility or fecundity; effects on sperm production, maturation, transport, morphology, or motility; malformations (hypospadias or cryptorchidism); alterations in size, weight, morphology, histology, or function of male reproductive organs (testis, epididymis, seminal vesicle, prostate, vas deferens, or gubernaculum); and changes in anogenital distance. **Secondary outcomes:** Indicators of male reproductive effects, including altered levels of endocrine or biochemical signaling molecules (fetal testosterone, fetal testis steroidogenic or cholesterologenic proteins, and insulin-like factor 3), receptors, or mRNAs; and changes in cell proliferation.

(2) Animal Study Question: Does in utero exposure to phthalates in nonhuman mammals cause male reproductive malformation?

TABLE 2 Animal PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Male nonhuman mammals without restriction based on species or age
Exposure	In utero exposure to one or more of the following phthalates or the corresponding monoester metabolite: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), and/or dipentyl phthalate (CAS no. 131-18-0). No restrictions based on route of exposure, based on administered dose or concentration, or biomonitoring data (e.g., urine, blood, or other specimen measurements).

(Continued)

TABLE 2 Continued

Element	Evidence
Comparators	Male nonhuman mammal populations exposed to different doses of the selected phthalates or vehicle-only treatment
Outcomes	**Primary outcomes:** Male reproductive effects, including alterations in fertility or fecundity; effects on sperm production, maturation, transport, morphology, or motility; malformations (hypospadias or cryptorchidism) or alterations in size, weight, morphology, histology, or function of male reproductive organs (testis, epididymis, seminal vesicle, prostate, vas deferens, or gubernaculum); and changes in anogenital distance. **Secondary outcomes:** Indicators of male reproductive effects, including altered levels of endocrine or biochemical signaling molecules (fetal testosterone, fetal testis steroidogenic or cholesterologenic proteins, and insulin-like factor 3), receptors, or mRNAs; and changes in cell proliferation.

PECO Statements for TCDD

(3) Human Study Question: Is developmental exposure to TCDD in humans associated with male reproductive effects?

TABLE 3 Human PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Males without restriction based on age
Exposure	Developmental exposure to TCDD (CAS no. 1746-01-6), with no restrictions based on route of exposure, based on biomonitoring data (e.g., urine, blood, or other specimens), environmental measures (e.g., air or water concentrations), or indirect measures (e.g., job title). To be considered "developmental" the exposure occurred during any of the following: pre-conception for one or both parents, prenatal to the pregnant female and/or directly to the fetus, or postnatal until sexual maturation.
Comparators	Populations exposed at lower levels of TCDD
Outcomes	**Primary outcomes:** Male reproductive effects, including alterations in fertility or fecundity; effects on sperm production, maturation, transport, morphology, or motility; malformations (hypospadias or cryptorchidism) or alterations in size, weight, morphology, histology, or function of male reproductive organs (testis, epididymis, seminal vesicle, prostate, vas deferens, or gubernaculum); altered age at puberty; and changes in anogenital distance. **Secondary outcomes:** Indicators of male reproductive effects, including altered levels of endocrine or biochemical signaling molecules (testosterone, luteinizing hormone, and insulin-like growth factor-1), receptors, or mRNAs.

(4) Animal Study Question: Does developmental exposure to TCDD in nonhuman mammals cause male reproductive effects?

TABLE 4 Animal PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Male nonhuman mammals without restriction based on species or age (including experimental or wildlife models)
Exposure	Developmental exposure to TCDD (CAS no. 1746-01-6), with no restrictions based on route of exposure, based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimen measurements), or environmental measurements (e.g., air or water concentrations). To be considered "developmental" the exposure occurred during any of the following: pre-conception for one or both parents, prenatal to the pregnant female and/or directly to the fetus, or postnatal until sexual maturation.
Comparators	Male nonhuman mammalian populations exposed to vehicle-only treatment in experimental studies or lower levels of TCDD in wildlife studies
Outcomes	**Primary outcomes:** Male reproductive effects, including alterations in fertility; effects on sperm production, maturation, transport, morphology, or motility; malformations (hypospadias or cryptorchidism) or alterations in size, weight, morphology, histology, or function of male reproductive organs (testis, epididymis, seminal vesicle, prostate, vas deferens, or gubernaculum); altered age at puberty; changes in anogenital distance; nipple or areola retention; and alterations in male-associated reproductive behaviors. **Secondary outcomes:** Indicators of male reproductive effects, including altered levels of endocrine or biochemical signaling molecules (testosterone, luteinizing hormone, insulin-like growth factor-1), receptors, or mRNAs.

PECO Statements for Bisphenol A

(5) Human Study Question: Is exposure to bisphenol A in humans associated with female reproductive effects?

TABLE 5 Human PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Females without restriction based on age
Exposure	Exposure to bisphenol A (CAS no. 80-05-7), with no restrictions based on route of exposure, based on exposure media (e.g., food or consumer product concentrations), biomonitoring data (e.g., urine, blood, or other specimen measurements), or indirect measures (e.g., job title).
Comparators	Populations exposed at lower levels of bisphenol A
Outcomes	**Primary outcomes:** Female reproductive effects, including alterations in fertility or fecundity (time-to-pregnancy and spontaneous abortion); alterations in ovulation or reproductive cyclicity; alterations in size, weight, morphology, histology, or function of female reproductive organs (ovaries, fallopian tubes, uterus, vagina, and mammary gland); altered age at puberty; adverse effects on lactation; premature reproductive senescence; changes in anogenital distance; changes in timing of breast development; and alterations in pubic hair development.

(Continued)

TABLE 5 Continued

Element	Evidence
Outcomes	**Secondary outcomes:** Indicators of female reproductive effects, including altered levels of endocrine or biochemical signaling molecules (androstenedione, dehydroepiandrosterone sulfate, estradiol, estrone, insulin-like growth factor-1, luteinizing hormone, sex hormone-binding globulin, and testosterone), receptors, or mRNAs; and changes in cell proliferation.

(6) Animal Study Question: Does exposure to bisphenol A in nonhuman mammals cause female reproductive effects?

TABLE 6 Animal PECO (Population, Exposure, Comparator, and Outcome) Statement

Element	Evidence
Population	Female nonhuman mammals without restriction based on species or age
Exposure	Exposure to bisphenol A (CAS no. 80-05-7), with no restrictions based on route of exposure, based on administered dose or concentration, or biomonitoring data (e.g., urine, blood, or other specimen measurements).
Comparators	Female nonhuman mammalian populations exposed to vehicle-only treatment in experimental studies or lower levels of bisphenol A found in background populations.
Outcomes	**Primary outcomes:** Female reproductive effects, including alterations in fertility or fecundity (time-to-pregnancy, spontaneous abortion, fetal loss, resorptions, and litter size); alterations in ovulation or reproductive cyclicity; alterations in size, weight, morphology, histology, or function of female reproductive organs (ovaries, fallopian tubes, uterus, vagina, and mammary gland); altered age at puberty; adverse effects on lactation; premature reproductive senescence; female-associated reproductive behaviors; and altered mammary gland development. **Secondary outcomes:** Indicators of female reproductive effects, including altered levels of endocrine or biochemical signaling molecules (androstenedione, dehydroepiandrosterone sulfate, estradiol, estrone, insulin-like growth factor-1, luteinizing hormone, sex hormone-binding globulin, and testosterone), receptors, or mRNAs; and changes in cell proliferation.

Questions for Each Panel:

1. Has the committee framed questions that can be addressed using systematic review methods? What changes would you suggest for the research questions, such as narrowing or widening the scope of each question?

2. Are the exposures adequately defined for each research question?
   a. Should any additional measures be added?
   b. Should any be modified or removed?
   c. Are there critical windows of susceptibility that should be considered in defining the exposures for each research question?
   d. What issues, such as toxicokinetics or analytical artifacts, should be considered in evaluating the internal and external validity of different exposure metrics?
   e. Are there confounding co-exposures that should be considered?
3. Please comment on the appropriateness of the selected comparators.

4. Are the primary or secondary outcomes appropriate?

a. Are there additional primary or secondary outcomes that are plausibly caused by altered endocrine function? If so, what are they?
b. In general, the primary outcomes are indicators of clinical effects that would be considered adverse whereas the secondary outcomes are considered surrogate measures (e.g., laboratory tests) that may be less predictive of adverse changes (e.g., upstream indicators).
    i. Does the panel agree with the way the committee identified primary and secondary outcomes?
    ii. Are there effects listed as primary health outcomes that should be considered secondary outcomes? If so, please provide an explanation and support for making the change.
    iii. Are there effects listed as secondary health outcomes that should be considered primary outcomes? If so, please provide an explanation and support for making the change.
    iv. Are there effects listed that should <u>not</u> be considered by the committee? If so, please provide an explanation and support for making the change.
    v. Are there effects not listed that should be considered by the committee? If so, please provide an explanation and support for making the change.

5. Are the animal populations relevant for the committee's statement of task?
    a. Are there major toxicokinetic or toxicodynamic differences across mammalian species or strains that should be considered in evaluating potential effects of these chemicals? If so, please provide key considerations or differences that should be considered, and to which outcomes such differences are applicable.
    b. Are there nonmammalian models that should be considered equal to or as reliable as mammalian models (e.g., rodents or nonhuman primates) for evaluating potential reproductive effects in humans?

6. Are there any particular study designs or characteristics that would make a study more or less valid or reliable?

7. Is the database sufficient to conduct a systematic review to address the PECO question?

Appendix C

Supporting Materials for the Phthalate (Animal) Systematic Review

SECTION C-1

PHTHALATE (ANIMAL) SYSTEMATIC REVIEW PROTOCOL

August 3, 2016
(Modified on September 15, 2016—See Section C-1f)

BACKGROUND AND INTRODUCTION

Phthalates are high production volume chemicals used primarily as plasticizers in many industrial and consumer products. As a result of their ubiquitous use, there is documented widespread human exposure to them. Because the developing organism has been shown to be particularly vulnerable to endocrine-disrupting chemicals, such as phthalates, the committee decided to focus on studies of in utero exposure. Ortho-phthalates have been linked to effects on male reproductive-tract development after in utero exposure in animal studies.

OBJECTIVE AND SPECIFIC AIMS

Review Question

The overall objective of this systematic review is to answer the question what is the effect of in utero exposure to ortho-phthalates on anogenital distance, hypospadias, or fetal testosterone in nonhuman male mammals?

The specific aims of the review are to:

- Identify literature reporting the effects of in utero phthalate exposure on male anogenital distance, hypospadias, or fetal testosterone in nonhuman mammals.
- Extract data on male effects of in utero phthalate exposure on anogenital distance, hypospadias, or fetal testosterone from relevant studies.
- Assess the internal validity (risk of bias) of individual studies.
- Summarize the extent of evidence available.
- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration, such as study-design heterogeneity.
- Rate the confidence in the body of evidence for studies in nonhuman mammals according to one of five statements: (1) high, (2) moderate, (3) low, (4) very low/no evidence available, or (5) evidence of lack of effects on male reproductive-tract development.

PECO Statement

A PECO (Population, Exposure, Comparator, and Outcome) statement was developed by the review team as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the review question for the systematic review.

Population: Nonhuman male mammals

Exposure:

- In utero exposure to any of the following ortho-phthalates or the corresponding monoester or oxidative metabolites: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or dipentyl phthalate (CAS no. 131-18-0).
- Oral route of exposure.

Comparator: Male nonhuman mammals exposed in utero to different doses of phthalates or vehicle-only treatment.

Outcomes:

- Anogenital distance (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:
  - Anogenital index (AGI): AGD measurement divided by body weight or by the cube root of body weight
  - Anoscrotal distance (ASD): the measured distance between the anus and base of the scrotum
  - Anopenile distance (APD): the measured distance from the anus to the base of the penis
- Hypospadias (incidence and severity/grade)
- Fetal testosterone concentration (e.g., measured from testes, serum, or plasma taken in utero)

METHODS

Problem Formulation and Protocol Development

The review question and specific aims were developed and refined through a series of problem formulation steps. The committee considered review articles on endocrine disruptors in surveying the types of chemicals that might make good case examples and held a workshop to explore potential case examples, including phthalates. The committee sought an example of a chemical for which the human and the animal evidence on effects appears to be associated with different exposure levels of that chemical and due to perturbation of the estrogen or androgen hormone system. Phthalates appear to fit this case criterion, and positive feedback was received at the committee's workshop.

Alterations in male reproductive-tract development are the most sensitive effects from exposure to phthalates (NRC 2008). Because the period during in utero sexual differentiation (i.e., the masculinization programming window) is the most sensitive life stage, the exposure period of interest for the systematic

review is in utero. Animal studies have illustrated a spectrum of effects in male reproductive development after in utero exposure to phthalates, including under developed or absent reproductive organs, malformed external genitalia (hypospadias), undescended testicles (cryptorchidism), decreased AGD, retained nipples, and decreased sperm production (NRC 2008). The systematic review will focus on end points reflecting androgen-dependent adverse effects (AGD and hypospadias), an adverse effect that occurs at relatively low doses (AGD), and a key event in the adverse outcome pathway leading to reduced AGD and hypospadias (fetal testosterone).

Consideration was given to including cryptorchidism as an end point, but the committee decided against it. The mode of action for phthalate-induced cryptorchidism involves reductions in INSL-3 levels in addition to androgen-dependent mechanisms. Important for the committee's charge, there are few, if any, human studies on dose-response relationships between phthalate exposure and cryptorchidism to compare to animal data. Furthermore, studies have shown that rats exposed to phthalates have similar sensitivity to decreased fetal testosterone and AGD as they do for decreased INSL-3, and that cryptorchidism is a less sensitive end point compared to reductions in AGD. Because the overall objective of the committee is to use this systematic review with the one being conducted on the human evidence to evaluate the coherence between effects and dose-response relationships, the committee judged that it would not be useful to include cryptorchidism in the systematic reviews on phthalates.

The protocol will be peer reviewed by subject-matter and systematic-review experts in accordance with standard report-review practices of the National Academies of Sciences, Engineering, and Medicine. The protocols will be revised in response to peer review comments and will subsequently be published as appendices to the committee's final report. The identity of the peer reviewers will remain anonymous to the committee until the publication of the final report, when their names and affiliations are disclosed in the Preface.

Committee and Staff

There are 11 committee members, supported by two staff members of the National Academies. The committee members were appointed in accordance with the standard policies and practices of the National Academies on the basis of their expertise in general toxicology, reproductive toxicology, developmental toxicology, endocrinology, neurotoxicology, epidemiology, risk assessment, biostatistics, and systematic-review methods. The membership of the committee and the staff was determined before the topic of the systematic review was selected. It was known, however, that each case study would be on an endocrine-disrupting chemical, so committee members who have relevant expertise were specifically recruited and appointed.

Review Team

The review team for this case study will be a subgroup of the committee (DD, KJ), two National Academies staff members (EM, SM), and an information specialist (JB). If a member of the review team was a coauthor of a study under review, that member will recuse himself or herself from the evaluation of the quality of that study.

The review team will be responsible for performing all aspects of the review, including conducting the literature searches; applying inclusion/exclusion criteria to screen studies; extracting data; assessing risk of bias for included studies; and analyzing and synthesizing data. The roles and responsibilities of the team members will be documented throughout the protocol. Throughout the course of its work, the review team will also engage other members of the committee to provide consultation as needed. The involvement of those individuals will be documented and acknowledged.

Biographical information on the review team is presented in Section C-1a.

Search Methods

Search for Existing Systematic Reviews

The review team will consider using existing systematic reviews to address or help to address its research question. English-language systematic reviews conducted within the last 3 years will be sought. The review team will incorporate prior reviews, update prior reviews, and/or use the reviews as part of its searching, depending on determination of their relevancy and quality (Whitlock et al. 2008). Current guidance on using existing systematic reviews will be used (Robinson et al. 2014, 2015, 2016).

Search

Recent, relevant, high-quality systematic reviews addressing the research question about phthalates and male reproductive-tract development will be searched. PubMed will be searched by adding the qualifier "systematic review"[ti] OR "meta-analysis"[pt] OR "meta-analysis"[ti] OR ("systematic"[ti] AND "review"[ti]) OR (systematic review [tiab] AND review [pt]) OR "meta synthesis"[ti] OR "meta synthesis"[ti] OR "integrative review"[tw] OR "integrative research review"[tw] OR "cochrane database syst rev"[ta] OR "evidence synthesis"[tiab] to the preliminary search strategy (see Section C-1b). Language and date restrictions will be applied (English language; published 2013 to present). The systematic review protocol registries PROSPERO (CRD) and CAMARADES will also be searched using key terms from the preliminary PubMed strategy.

Study Selection

Two team members (SM, EM) will independently screen search results, applying the following exclusion criteria:

- Not a systematic review.[1] The minimum criteria for a study to be considered a systematic review are
  - conduct of an explicit and adequate literature search,
  - application of predefined eligibility criteria,
  - consideration of the quality of included studies or risk of bias assessment, and
  - synthesis (or attempt at synthesis) of the findings, either qualitatively or quantitatively.
- Not in English.
- Search date prior to 2013.
- Does not match the research question or PECO elements.

For PubMed results, screening will be conducted first using abstracts and then at the full-text level. Results from PROSPERO and CAMARADES will be conducted at one level, using the information in the registry. Disagreements regarding eligibility will be resolved through discussion or, where necessary, by a third team member.

Assessment for Quality

Two investigators (KR, AR) will independently assess the risk of bias of eligible systematic reviews using ROBIS (Whiting et al. 2016). Disagreements in rating will be resolved through discussion or, where necessary, through consultation with a third team member. Systematic reviews rated as low quality will be excluded from further consideration at this stage.

[1]A systematic review "is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

Use of Existing Reviews

Eligible systematic reviews of high quality will be reviewed, considering date of search and match with the PECO statement as well as availability of data from the primary studies, how risk of bias was conducted, and other factors. Current reviews considered a good match will be used to address the research question. Reviews that are a good match but with search dates more than a year ago will be updated. If no relevant systematic reviews are found, an independent systematic review will be performed.

Literature Search for Independent Systematic Review

The review team will collaborate with an information specialist (JB) who has training, expertise, and familiarity with developing and performing systematic review literature searches. A variety of methods will be used to identify relevant data (see below). Literature searches will not be limited by publication date.

Online Databases

Electronic searches of the following three online databases will be performed using the search terms outlined in Section C-1b: PubMed, Embase, and Toxline. The search strategy and search terms will be developed by the information specialist (JB), who will implement the search for relevant studies.

Other Resources

Hand searching the reference lists of all the included studies after full-text review will be conducted using the same study selection process as used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

Study Selection

All search results will be imported or manually entered into EndNote (Version x7) reference management software. EndNote will be used to eliminate any duplicate citations before evaluating the eligibility of the citations.

Screening Process

References retrieved from the literature search will be screened for relevance and eligibility against the evidence selection criteria using DistillerSR (Evidence Partners; https://www.evidencepartners.com/). Screeners from the review team will be trained with an initial pilot phase on 25 studies undertaken to improve clarity of the evidence selection criteria and to improve accuracy and consistency among screeners. Screening forms are presented in Section C-1c.

Title and Abstract Screening

Each citation will be independently screened by two reviewers (SM, EM) to determine whether it meets the selection criteria for inclusion that reflect the PECO statement with some additional considerations as listed below. Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

The title/abstract screening form will be used to screen and EXCLUDE references if at least one of the following criteria is met:

1. No original data (e.g., review article, commentary, editorial)
2. Study does not include nonhuman mammals
3. Study does not report phthalate exposure
4. No relevant outcomes
5. Incomplete information (e.g., conference abstract, meeting poster)
6. Not in English and unable to determine eligibility
7. Other (explanation required)

The following types of records will be INCLUDED at the title/abstract level: any English-language study of male humans exposed to phthalates in utero.

Only English-language publications will be included because of time and resource constraints. There appears to be no indication that foreign-language publications would make a contribution that is distinct from what is found in the English-language literature.

Updated details to instructions and interpretations for title and abstract screening will be added to the Section C-1f to document the process of the review team during the screening process.

Full-Text Screening

Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers involved in title and abstract screening (SM, EM). Each reference will be screened in duplicate and independently. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

Citations will be EXCLUDED at the full-text level if at least one of the following criteria are met:

1. No original data (e.g., review article, commentary, editorial)
2. Study does not include nonhuman mammals
3. Study does not report experimental exposure to one or more of the phthalates listed in the PECO statement
4. Study does not report oral exposure to phthalates
5. Study does not quantify exposure to phthalates
6. Study does not include in utero exposure
7. Study does not assess or report anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or fetal testosterone concentrations
8. No comparator group (animals exposed to different doses of phthalates or vehicle-only treatment)
9. Not in English
10. Other reason (explanation required).

The reason for exclusion at the full-text-review stage will be annotated and reported in a study selection flow diagram in the final report (following PRISMA [Moher et al. 2009]). The reasons for exclusion will be documented from the list (1-9) above.

Citations will be INCLUDED if they meet the PECO statement criteria:

- Study includes nonhuman male mammals
- Study includes in utero exposure
- Study includes comparison with animals exposed to different doses of phthalates or vehicle-only treatment

- Study measures anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or fetal testosterone

Updated details to instructions and interpretations for full-text screening will be added to the Section C-1f to document the process of the review team during the screening process.

Data Extraction

Data will be collected and recorded (i.e., extracted) from included studies by one member of the review team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction will be resolved through discussion. The extracted data will be used to summarize study designs and findings and/or to conduct statistical analyses. Section C-1d presents the data extraction elements that will be used.

The review team will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The study extraction files will note whether an attempt was made to contact study authors by email for missing data considered important for evaluating key study findings (and whether or not a response was received).

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 review will include all publications on the study, select one publication to use as the primary publication, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up or, for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. The review will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, the review team will include a single instance of the data (and avoid more than one—that is, duplicate instances of the data).

Data extraction will be completed using the Health Assessment Workspace Collaborative (HAWC) software, an open source and freely available Web-based interface application, for visualization and warehousing.[2]

Risk of Bias (Quality) Assessment of Indiviual Studies

Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect. Internal validity or risk of bias will be assessed for individual studies using a tool developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) that outlines an approach to evaluating risk of bias for experimental animal studies. The risk of bias domains and questions for experimental animal studies are based on established guidance for experimental human studies (randomized clinical trials) (Higgins and Green 2011; Viswanathan et al. 2012, 2013; Sterne et al. 2014) and recent tools for animal studies (Hooijmans et al. 2014; Koustas et al. 2014). The riskof bias tool includes a common set of questions (Section C-1e) that are answered based on the specific details of individual studies to develop risk of bias ratings (using the four options: definitely low risk of bias; probably low risk of bias; probably high risk of bias; or definitely high risk of bias).

[2]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

Information or study procedures that were not reported are assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Study design determines the subset of questions that should be used to assess risk of bias for an individual study (see Table C1-1).

Studies are independently assessed by two assessors (DD, KJ) who answer all applicable risk of bias questions with one of four options (see Table C1-2) following prespecified criteria detailed in Section C-1e. The criteria describe aspects of study design, conduct, and reporting required to reach risk of bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific type of human study designs. Risk of bias will be assessed at the outcome level because study design or method specifics may increase the risk of bias for some outcomes and not others within the same study. Information or study procedures that were not reported are assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Authors will be queried by email to obtain missing information, and responses received will be used to update riskof bias ratings.

Assessors will be trained using the criteria in an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk of bias assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk of bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk of bias criteria.

After assessors have independently made risk of bias determinations for a study across all risk of bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered and resolved with the review team. The final riskof bias rating for each question will be recorded along with a statement of the basis for that rating.

Data Analysis and Evidence Synthesis

The review team will qualitatively synthesize the body of evidence for each outcome and, where appropriate, a meta-analysis will be performed. If a meta-analysis is performed, summaries of main characteristics for each included study will be compiled and reviewed by two team members to determine comparability between studies, to identify data transformations necessary to ensure comparability, and to determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

- Experimental design (e.g., acute, chronic, multigenerational)
- Animal model used (e.g., species, strain, sex, genetic background)
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status)
- Developmental stage of animals at treatment and outcome assessment
- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

The review team expects to require input from subject-matter experts to help assess the heterogeneity of the studies. Subgroup analyses to examine the extent to which risk of bias contributes to heterogeneity will be performed. If there is evidence of species differences, the review team will consider stratifying

by species and performing separate meta-analyses by species. Situations where it may not be appropriate to include a study are when data on exposure or outcome are too different to be combined or other circumstances that may indicate that averaging study results would not produce meaningful results. When considering outcome measures for conducting meta-analyses, benchmark dose (BMD) estimates (and their associated confidence intervals) with a benchmark response (BMR) set to a common percent of control (for continuous outcomes) or extra risk (for dichotomous outcomes) are preferred. A secondary alternative, when there are more than two groups, is the conduct of BMD modeling and the use of the derived BMD estimates. Meta-analyses are not possible with lowest-observed-adverse-effect levels or no-observed-adverse-effect levels, since no confidence interval can be derived for these measures.

If a meta-analysis is conducted, a random effects model will be used for the analysis. Heterogeneity will be assessed using the I-squared statistic. Interpretation of I-squared will be based on the Cochrane Handbook: 0% to 40% (might not be important); 30% to 60% (may represent moderate heterogeneity); 50% to 90% (may represent substantial heterogeneity); and 75% to 100% (considerable heterogeneity). Additionally, as described in the Cochrane Handbook, for the last three categories, the importance of the I-squared will be interpreted considering not only the magnitude of effects but also the strength of the evidence (90% two-tailed confidence interval).

The review team will also perform sensitivity analyses on the exclusion of individual studies in succession.

If sufficient studies are available, subgroup analyses will be performed based on the following characteristics described above: experimental design, animal model used (e.g., species and/or strain), age of animals, and developmental stage of animals at treatment and outcome assessment.

In the event that these proposed methods for data analysis are altered to tailor to the evidence base from included studies, the protocol will be amended accordingly, and the reasons for change will be justified in the documentation.

Confidence Rating: Assessment of the Body of Evidence

The quality of evidence for each male reproductive outcome will be evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in NTP (2015, see Step 5). In brief, available studies on a particular outcome are initially grouped by key study-design features, and each grouping of studies is given an initial confidence rating by those features.

The initial rating is downgraded for factors that decrease confidence in the results, including

- risk of bias
- unexplained inconsistency
- indirectness or lack of applicability
- imprecision
- publication bias

The initial rating is upgraded for factors that increase confidence in the results, including

- large magnitude of effect
- dose response
- consistency across study designs/populations/animal models or species
- consideration of residual confounding
- other factors that increase confidence in the association or effect (e.g., particularly rare outcomes)

TABLE C1-1 OHAT Risk of Bias Tool

Risk-of-Bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.
**Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, non-randomized experimental studies)
***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).
SOURCE: NTP (2015, p. 37).

TABLE C1-2 Answers to the Risk of Bias Questions

++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

SOURCE: NTP (2015, p. 36).

The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt et al. 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney et al. 2014).

Confidence ratings are independently assessed by members of the review team, and discrepancies will be resolved by consensus and consultation with technical advisors as needed. Confidence ratings will be summarized in evidence profile tables.

REFERENCES

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

Higgins, J., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available: http://handbook.cocharne.org [accessed May 6, 2016].

Hooijmans, C.R., M.M. Rovers, R.B. de Vries, M. Leenars, M. Ritskes-Hoitinga, and M.W. Langendam. 2014. SYRCLE's risk of bias tool for animal studies. BMC Med. Res. Method. 14:43.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Koustas, E., J. Lam, P. Sutton, P.I. Johnson, D.S. Atchley, S. Sen, K.A. Robinson, D.A. Axelrad, and T.J. Woodruff. 2104. The Navigation Guide—evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122(10):1015-1027.

Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. J. Clin. Epidemiol. 62(10):1006-1012.

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-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Robinson, K.A., E.P. Whitlock, M.E. O'Neil, J.K. Anderson, L. Hartling, D.M. Dryden, M. Butler, S.J. Newberry, M. McPheeters, N.D. Berkman, J.S. Lin, and S. Chang S. 2014. Integration of Existing Systematic Reviews. Research white paper. AHRQ Publication No. 14-EHC016-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK216379/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2015. Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 15-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279904/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2016. Twelve recommendations for integrating existing systematic reviews into new reviews: EPC guidance. J. Clin. Epidemiol. 70:38-44.

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. Environ. Health Perspect. 122(7):711-718.

Sterne, J.A.C., J.P.T. Higgins, and B.C. Reeves. 2014. ACROBAT-NRSI: A Cochrane risk of Bias Assessment Tool for Non-randomized Studies of Interventions. Version 1.0.0. September 24, 2014 [online]. Available: www.riskofbias.info [accessed May 6, 2016].

Viswanathan, M., M. Ansari, N.D. Berkman, S. Chang, L. Hartling, L.M. McPheeters, P.L. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze, and J.R. Treadwell. 2012. Assessing the Risk of Bias of Individual Studies when Comparing Medical Interventions. AHRQ Publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available www.effectivehealthcare.ahrq.gov/ [accessed May 6, 2016].

Viswanathan, M., N.D. Berkman, D.M. Dryden, and L. Hartling. 2013. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank. Methods Research Report. AHRQ Publication No. 13-EHC106-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available www.effectivehealthcare.ahrq.gov/reports/final.cfm [accessed May 6, 2016].

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

Whitlock, E.P., J.S. Lin, R. Chou, P. Shekelle, and K.A. Robinson. 2008. Using existing systematic reviews in complex systematic reviews. Ann. Intern. Med. 148(10):776-782.

SECTION C-1a

REVIEW TEAM BIOGRAPHICAL INFORMATION

Jaime F. Blanck is a clinical informationist at the Welch Medical Library at Johns Hopkins University. She creates and implements systematic review search strategies across multiple databases and provides comprehensive reference, research, and information services to multiple departments within the School of Medicine. She received an MLIS from the University of Pittsburgh and an MPA from the University of Baltimore.

David C. Dorman is a professor of toxicology in the Department of Molecular Biosciences of North Carolina State University. The primary objective of his research is to provide a refined understanding of chemically induced neurotoxicity in laboratory animals that will lead to improved assessment of potential toxicity in humans. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has chaired or served on several NRC committees, including the Committee on Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decisions, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He has served on other advisory boards for the US Navy, NASA, and USDA, and is currently a member of the NTP's Board of Scientific Counselors. He is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Sciences. He received his DVM from Colorado State University. He completed a combined PhD and residency program in toxicology at the University of Illinois at Urbana-Champaign, and is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.

Kamin J. Johnson is a lead scientist for The Dow Chemical Company's Toxicology and Environmental Research and Consulting function, responsible for the scientific conduct and interpretation of developmental and reproductive toxicology studies. He has served on study sections of the National Institutes of Health reviewing reproductive toxicology grants, and he was a counselor for the Reproductive and Toxicology Specialty Section of the Society of Toxicology. His research interests are in the molecular and cellular biology of fetal and postnatal testis function, as well as mechanisms of testicular toxicants. Dr. Johnson received a PhD in molecular biology, cell biology, and biochemistry from Brown University.

Ellen Mantus is a scholar and director of risk assessment on the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine, with more than 20 years of experience in the fields of toxicology and risk assessment. She has served as the study director on numerous projects, including ones that have assessed the health implications of various chemical exposures; developed strategies for applying modern scientific approaches in toxicology and risk assessment; provided guidance to federal agencies on risk-based decision making; and evaluated barriers to deployment of electric vehicles and associated charging infrastructure. Before joining the National Academies, Dr. Mantus was a project manager with ICF Consulting where she served as a primary reviewer for numerous toxicological studies and provided risk assessment and regulatory support on a wide array of projects. Dr. Mantus received a PhD in chemistry from Cornell University.

Susan Martel is a senior program officer in the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine. She has 20 years of experience in supporting toxicology and risk assessment projects for the US Environmental Protection Agency, the US Department of Defense, and the National Aeronautics and Space Administration. Recent projects include working with committees evaluating the toxicological effect of arsenic, developing exposure guidelines for use on spacecraft, and assessing pesticide risks-assessment practices. Before joining the National

Academies, she was the administrator of the Registry for Toxicology Pathology for Animals at the American Registry of Pathology. She received a BA in biology from Skidmore College.

Andrew A. Rooney is deputy director of the Office of Health Assessment and Translation (OHAT) in the National Toxicology Program at the National Institute of Environmental Health Sciences. He has been developing risk assessment methods and guidance throughout his professional career and is a principal author of the 2012 WHO/IPCS Guidance for Immunotoxicity Risk Assessment for Chemicals. Most recently, he has been working on emerging issues in toxicology and environmental health, including methods to address study quality in terms of risk of bias for human, animal, and mechanistic studies and adaptation of systematic review methods for addressing environmental health questions. He led the team that developed the OHAT approach to systematic review. Dr. Rooney has an MS and a PhD in zoology from the University of Florida.

SECTION C-1b

LITERATURE SEARCH STRATEGY

The review team will employ a multi-method process to identify all potentially relevant studies as detailed below.

Electronic Searches

PubMed

A search string employing medical subject heading (MeSH) terms and keyword synonyms will be developed. The PubMed search strategy will be considered the primary search strategy and will provide the basis of the other electronic search strategies. To assist in compiling these terms, the review team will conduct a text analysis of 25 articles known to the authors. These articles were selected because they represent both American and non-American publications and will help identify spelling variants. The search strategies will address each of the following concepts:

- *Phthalates*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the following phthalates: the CAS numbers to these 11 phthalates: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or dipentyl phthalate (CAS no. 131-18-0). The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. CAS registry numbers for each phthalate substance will also be included in the list of search terms. All MeSH terms, Supplementary Concept terms, keyword synonyms, and CAS registry numbers will be searched together as one concept using the Boolean operator "OR."
- *Exposure*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the *exposure* concept. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. All MeSH terms and keyword synonyms will be searched together as one concept using the Boolean operator "OR."
- *Animal studies*—The review team will adapt the search filter published in *Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Laboratory Animals. 2010;44(3):170-175* to eliminate nonmammalian animals*. doi:10.1258/la.2010.009117.*

Each of the above concepts will be searched together using the Boolean operator "AND." There will not be limitations on date of publication, language, or publication type. All citation records will be exported to EndNote. Additional citations identified through the search processes identified below will also be exported to the project EndNote library. Duplicates will be removed from the citation library using the "Find Duplicates" tool in EndNote as well as a manual review of citations by the project librarian to identify any duplicates not found during the automated process. The number of citations found in each database will be recorded, as well as the number of duplicates and final tally of unique citations. The final library of citations will be uploaded to the Health Assessment Workspace Collaboration Web-based tool (www.hawcproject.org) for systematic reviews where they will be reviewed by the team.

Embase

The controlled vocabulary database Emtree is used by Embase. For each MeSH term identified through the process above, Emtree will be searched for the appropriate corresponding term. Additional keywords will be identified using the list of synonyms from each Emtree record and added to the keywords from the MeSH records. The review team will substitute the animal study search filter used in the PubMed search with the comparable Embase filter published in *De Vries RBM, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. A search filter for increasing the retrieval of animal studies in Embase.Laboratory Animals. 2011;45(4):268-270. doi:10.1258/la.2011.011056.* This version of the animal filter will also be adapted to remove all nonmammalian animals.

Toxline

The review team will develop the Toxline search strategy by removing any database specific formatting from the PubMed search strategy to create a keyword-only search (Toxline does not employ a controlled vocabulary).

Search Strategies

PubMed

(“butylbenzyl phthalate” [Supplementary Concept] OR “Dibutyl Phthalate”[Mesh] OR “diethyl phthalate” [Supplementary Concept] OR “Diethylhexyl Phthalate”[Mesh] OR “diisobutyl phthalate” [Supplementary Concept] OR “diisononyl phthalate” [Supplementary Concept] OR “diisooctyl phthalate” [Supplementary Concept] OR “dimethyl phthalate” [Supplementary Concept] OR “di-n-octyl phthalate” [Supplementary Concept] OR “benzylbutyl phthalate”[tw] OR “benzyl butyl phthalate”[tw] OR “butyl benzyl phthalate”[tw] OR “butylbenzyl phthalate”[tw] OR “butylbenzylphthalate”[tw] OR “phthalic acid butyl benzyl ester”[tw] OR “butyl-benzyl-phthalate”[tw] OR “BBzP”[tw] OR “BzBP”[tw] OR “BBPHT”[tw] OR “85-68-7”[tw] OR “Dibutyl Phthalate”[tw] OR “Di-n-Butyl Phthalate”[tw] OR “Di n Butyl Phthalate”[tw] OR “Butyl Phthalate”[tw] OR “d n butyl phthalate”[tw] OR “dbp”[tw] OR “di n butyl phthalate”[tw] OR “dibutyl phthalate”[tw] OR “dibutylphthalate”[tw] OR “phthalic acid di n butyl este”[tw] OR “84-74-2”[tw] OR “phthalic acid diethyl ester”[tw] OR “diethyl phthalate”[tw] OR “diethylphthalate”[tw] OR “ethyl phthalate”[tw] OR “di-ethyl phthalate”[tw] OR “DEP”[tw] OR “84-66-2”[tw] OR “bis (2 ethylhexyl) phthalate”[tw] OR “bis (2 ethylhexyl) phthalate”[tw] OR “bis (2 ethylhexyl) phthalate”[tw] OR “bis (2 ethylhexylphthalate)”[tw] OR “Bis(2-ethylhexyl)phthalate”[tw] OR “DEHP”[tw] OR “di (2 ethylhexyl) phthalate”[tw] OR “di 2 ethylhexyl phthalate”[tw] OR “di 2 ethylhexylphthalate”[tw] OR “Di-2-Ethylhexylphthalate”[tw] OR “diethylhexyl phthalate”[tw] OR “Di-octyl Phthalate”[tw] OR “octoil”[tw] OR “phthalic acid di 2 ethylhexyl ester”[tw] OR “phthalic acid diethylhexyl ester”[tw] OR “117-81-7”[tw] OR “di-iso-butyl phthalate”[tw] OR “DiBP”[tw] OR “84-69-5”[tw] OR “di-isononylphthalate”[tw] OR “ENJ 2065”[tw] OR “ENJ-2065”[tw] OR “di-isononyl phthalate”[tw] OR “di-iso-nonyl phthalate”[tw] OR “DINP”[tw] OR “28553-12-0”[tw] OR “Diisooctylphthalate”[tw] OR “27554-26-3”[tw] OR “diamyl phthalate”[tw] OR “dipentyl phthalate”[tw] OR “phthalic acid dipentyl ester”[tw] OR “dipentyl benzene-1,2-dicarboxylate”[tw] OR “di-n-pentyl phthalate”[tw] OR “131-18-0”[tw] OR “Dimethyl phthalate”[tw] OR “Dimethylphthalate”[tw] OR “Avolin”[tw] OR “Citrola”[tw] OR “Dmp”[tw] OR “dmp30”[tw] OR “fermine”[tw] OR “methyl phthalate”[tw] OR “mipax”[tw] OR “mugia”[tw] OR “palatinol m”[tw] OR “sketofax”[tw] OR “131-11-3”[tw] OR “Di-n-octyl phthalate”[tw] OR “di n octyl phthalate”[tw] OR “di n octylphthalate”[tw] OR “dioctyl phthalate”[tw] OR “dioctylphthalate”[tw] OR “di(n-octyl)phthalate”[tw] OR “phthalic acid di n octyl ester”[tw] OR “DNOP”[tw] OR “117-84-0”[tw]) AND (“Maternal Exposure”[Mesh] OR “Environmental Exposure”[Mesh:NoExp] OR “Prenatal Exposure Delayed Effects”[Mesh] OR “Exposure”[tw] OR “Exposed”[tw] OR “exposures”[tw] OR “exposing”[tw] AND (“Genital Diseases, Male”[Mesh] OR

"Genitalia, Male"[Mesh] OR "Testosterone"[Mesh:NoExp] OR "Androgens"[Mesh] OR "Anogenital"[tw] OR "AGD"[tw] OR "AGI"[tw] OR "ASD"[tw] OR "APD"[tw] OR "Urogenital"[tw] OR "Penile"[tw] OR "penis"[tw] OR "Anoscrotal"[tw] OR "Anopenile"[tw] OR "anorectal"[tw] OR "Testosterone"[tw] OR "androgen"[tw] OR "androgens"[tw] OR "Hypospadias"[tw] OR "hypospadia"[tw] OR "Testis"[tw] OR "testes"[tw] OR (("Anorectal"[tw] OR "genital"[tw] OR "genitals"[tw] OR "testes"[tw] OR "rectum"[tw]) AND ("malformation"[tw] OR "malformations"[tw] OR "development"[tw] OR "abnormalities"[tw] OR "abnormality"[tw] OR "dysplasia"[tw])) OR ("Male"[tw] and ("genital"[tw] OR "genitals"[tw] OR "genitalia"[tw])) OR ("Anus"[tw] AND ("genital"[tw] OR "genitals"[tw] OR "genitalia"[tw]))) AND ("animal experimentation"[MeSH Terms] OR "models, animal"[MeSH Terms] OR "invertebrates"[MeSH Terms] OR "Animals"[Mesh:noexp] OR "animal population groups"[MeSH Terms] OR "mammals"[MeSH Terms:noexp] OR "primates"[MeSH Terms:noexp] OR "artiodactyla"[MeSH Terms] OR "carnivora"[MeSH Terms] OR "cetacea"[MeSH Terms] OR "chiroptera"[MeSH Terms] OR "elephants"[MeSH Terms] OR "hyraxes"[MeSH Terms] OR "insectivora"[MeSH Terms] OR "lagomorpha"[MeSH Terms] OR "marsupialia"[MeSH Terms] OR "monotremata"[MeSH Terms] OR "perissodactyla"[MeSH Terms] OR "rodentia"[MeSH Terms] OR "scandentia"[MeSH Terms] OR "sirenia"[MeSH Terms] OR "xenarthra"[MeSH Terms] OR "haplorhini"[MeSH Terms:noexp] OR "strepsirhini"[MeSH Terms] OR "platyrrhini"[MeSH Terms] OR "tarsii"[MeSH Terms] OR "catarrhini"[MeSH Terms:noexp] OR "cercopithecidae"[MeSH Terms] OR "hylobatidae"[MeSH Terms] OR "hominidae"[MeSH Terms:noexp] OR "gorilla gorilla"[MeSH Terms] OR "pan paniscus"[MeSH Terms] OR "pan troglodytes"[MeSH Terms] OR "pongo pygmaeus"[MeSH Terms] animals[tiab] OR animal[tiab] OR mice[tiab] OR mus[tiab] OR mouse[tiab] OR murine[tiab] OR woodmouse[tiab] OR rats[tiab] OR rat[tiab] OR murinae[tiab] OR muridae[tiab] OR cottonrat[tiab] OR cottonrats[tiab] OR hamster[tiab] OR hamsters[tiab] OR cricetinae[tiab] OR rodentia[tiab] OR rodent[tiab] OR rodents[tiab] OR pigs[tiab] OR pig[tiab] OR swine[tiab] OR swines[tiab] OR piglets[tiab] OR piglet[tiab] OR boar[tiab] OR boars[tiab] OR "sus scrofa"[tiab] OR ferrets[tiab] OR ferret[tiab] OR polecat[tiab] OR polecats[tiab] OR "mustela putorius"[tiab] OR "guinea pigs"[Tiab] OR "guinea pig"[Tiab] OR cavia[Tiab] OR callithrix[Tiab] OR marmoset[Tiab] OR marmosets[Tiab] OR cebuella[Tiab] OR hapale[Tiab] OR octodon[Tiab] OR chinchilla[Tiab] OR chinchillas[Tiab] OR gerbillinae[Tiab] OR gerbil[Tiab] OR gerbils[Tiab] OR jird[Tiab] OR jirds[Tiab] OR merione[Tiab] OR meriones[Tiab] OR rabbits[Tiab] OR rabbit[Tiab] OR hares[Tiab] OR hare[Tiab] OR cats[Tiab] OR cat[Tiab] OR felis[Tiab] OR dogs[Tiab] OR dog[Tiab] OR canine[Tiab] OR canines[Tiab] OR canis[Tiab] OR sheep[Tiab] OR sheeps[Tiab] OR mouflon[Tiab] OR mouflons[Tiab] OR ovis[Tiab] OR goats[Tiab] OR goat[Tiab] OR capra[Tiab] OR capras[Tiab] OR rupicapra[Tiab] OR chamois[Tiab] OR haplorhini[Tiab] OR monkey[Tiab] OR monkeys[Tiab] OR anthropoidea[Tiab] OR anthropoids[Tiab] OR saguinus[Tiab] OR tamarin[Tiab] OR tamarins[Tiab] OR leontopithecus[Tiab] OR hominidae[Tiab] OR ape[Tiab] OR apes[Tiab] OR pan[Tiab] OR paniscus[Tiab] OR "pan paniscus"[Tiab] OR bonobo[Tiab] OR bonobos[Tiab] OR "pan troglodytes"[Tiab] OR gibbon[Tiab] OR gibbons[Tiab] OR siamang[Tiab] OR siamangs[Tiab] OR nomascus[Tiab] OR symphalangus[Tiab] OR chimpanzee[Tiab] OR chimpanzees[Tiab] OR prosimians[Tiab] OR "bush baby"[Tiab] OR prosimian[Tiab] OR bush babies[Tiab] OR galagos[Tiab] OR galago[Tiab] OR pongidae[Tiab] OR gorilla[Tiab] OR gorillas[Tiab] OR pongo[Tiab] OR "pongo pygmaeus"[Tiab] OR orangutans[Tiab] OR lemur[Tiab] OR lemurs[Tiab] OR lemuridae[Tiab] OR horse[Tiab] OR horses[Tiab] OR pongo[Tiab] OR equus[Tiab] OR cow[Tiab] OR calf[Tiab] OR bull[Tiab] OR chicken[Tiab] OR chickens[Tiab] OR squirrel[Tiab] OR squirrels[Tiab] OR chipmunk[Tiab] OR chipmunks[Tiab] OR suslik[Tiab] OR susliks[Tiab] OR vole[Tiab] OR voles[Tiab] OR lemming[Tiab] OR lemmings[Tiab] OR muskrat[Tiab] OR muskrats[Tiab] OR lemmus[Tiab] OR otter[Tiab] OR otters[Tiab] OR marten[Tiab] OR martens[Tiab] OR martes[Tiab] OR weasel[Tiab] OR badger[Tiab] OR badgers[Tiab] OR ermine[Tiab] OR mink[Tiab] OR minks[Tiab] OR sable[Tiab] OR sables[Tiab] OR gulo[Tiab] OR gulos[Tiab] OR wolverine[Tiab] OR wolverines[Tiab] OR minks[Tiab] OR mustela[Tiab] OR llama[Tiab] OR llamas[Tiab] OR alpaca[Tiab] OR alpacas[Tiab] OR camelid[Tiab] OR camelids[Tiab] OR guanaco[Tiab] OR guanacos[Tiab] OR chiroptera[Tiab] OR chiropteras[Tiab] OR bat[Tiab] OR bats[Tiab] OR fox[Tiab] OR foxes[Tiab] OR donkey[Tiab] OR donkeys[Tiab] OR mule[Tiab] OR mules[Tiab] OR zebra[Tiab] OR zebras[Tiab] OR

shrew[Tiab] OR shrews[Tiab] OR bison[Tiab] OR bisons[Tiab] OR buffalo[Tiab] OR buffaloes[Tiab] OR deer[Tiab] OR deers[Tiab] OR bear[Tiab] OR bears[Tiab] OR panda[Tiab] OR pandas[Tiab] OR "wild hog"[Tiab] OR "wild boar"[Tiab] OR fitchew[Tiab] OR fitch[Tiab] OR beaver[Tiab] OR beavers[Tiab] OR jerboa[Tiab] OR jerboas[Tiab] OR capybara[Tiab] OR capybaras[Tiab])

Embase

"phthalic acid benzyl butyl ester"/exp OR "phthalic acid dibutyl ester"/exp OR "phthalic acid diethyl ester"/exp OR "phthalic acid bis(2 ethylhexyl) ester"/exp OR "phthalic acid dimethyl ester"/exp OR "phthalic acid dioctyl ester"/exp OR "benzylbutyl phthalate" OR "benzyl butyl phthalate" OR "butyl benzyl phthalate" OR "butylbenzyl phthalate" OR "butylbenzylphthalate" OR "phthalic acid butyl benzyl ester" OR "butyl-benzyl-phthalate" OR "BBzP" OR "BzBP" OR "BBPHT" OR "85-68-7" OR "Dibutyl Phthalate" OR "Di-n-Butyl Phthalate" OR "Di n Butyl Phthalate" OR "Butyl Phthalate" OR "d n butyl phthalate" OR "dbp" OR "di n butyl phthalate" OR "dibutyl phthalate" OR "dibutylphthalate" OR "phthalic acid di n butyl este" OR "84-74-2" OR "phthalic acid diethyl ester" OR "diethyl phthalate" OR "diethylphthalate" OR "ethyl phthalate" OR "di-ethyl phthalate" OR "DEP" OR "84-66-2" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexylphthalate)" OR "Bis(2-ethylhexyl)phthalate" OR "DEHP" OR "di (2 ethylhexyl) phthalate" OR "di 2 ethylhexyl phthalate" OR "di 2 ethylhexylphthalate" OR "Di-2-Ethylhexylphthalate" OR "diethylhexyl phthalate" OR "Dioctyl Phthalate" OR "octoil" OR "phthalic acid di 2 ethylhexyl ester" OR "phthalic acid diethylhexyl ester" OR "117-81-7" OR "di-iso-butyl phthalate" OR "DiBP" OR "84-69-5" OR "di-isononylphthalate" OR "ENJ 2065" OR "ENJ-2065" OR "di-isononyl phthalate" OR "di-isononyl phthalate" OR "DINP" OR "28553-12-0" OR "Diisooctylphthalate" OR "27554-26-3" OR "diamyl phthalate" OR "dipentyl phthalate" OR "phthalic acid dipentyl ester" OR "dipentyl benzene-1,2-dicarboxylate" OR "di-n-pentyl phthalate" OR "131-18-0" OR "Dimethyl phthalate" OR "Dimethylphthalate" OR "Avolin" OR "Citrola" OR "Dmp" OR "dmp30" OR "fermine" OR "methyl phthalate" OR "mipax" OR "mugia" OR "palatinol m" OR "sketofax" OR "131-11-3" OR "Di-n-octyl phthalate" OR "di n octyl phthalate" OR "di n octylphthalate" OR "dioctyl phthalate" OR "dioctylphthalate" OR "di(n-octyl)phthalate" OR "phthalic acid di n octyl ester" OR "DNOP" OR "117-84-0" AND ('male genital system disease'/exp OR 'male genital system'/exp OR 'testosterone'/exp OR 'androgen'/de OR "Anogenital":ti,ab OR "AGD":ti,ab OR "AGI":ti,ab OR "ASD":ti,ab OR "APD":ti,ab OR "Urogenital":ti,ab OR "Penile":ti,ab OR "penis":ti,ab OR "Anoscrotal":ti,ab OR "Anopenile":ti,ab OR "anorectal":ti,ab OR "Testosterone":ti,ab OR "androgen":ti,ab OR "androgens":ti,ab OR "Hypospadias":ti,ab OR "hypospadia":ti,ab OR "Testis":ti,ab OR "testes":ti,ab OR (("Anorectal":ti,ab OR "genital":ti,ab OR "genitals":ti,ab OR "testes":ti,ab OR "rectum":ti,ab) AND ("malformation":ti,ab OR "malformations":ti,ab OR "development":ti,ab OR "abnormalities":ti,ab OR "abnormality":ti,ab OR "dysplasia":ti,ab)) OR ("Male":ti,ab and ("genital":ti,ab OR "genitals":ti,ab OR "genitalia":ti,ab)) OR ("Anus":ti,ab AND ("genital":ti,ab OR "genitals":ti,ab OR "genitalia":ti,ab))) AND ('prenatal exposure'/exp OR 'environmental exposure'/exp OR 'exposure' OR 'exposed' OR 'exposures' OR 'exposing' AND ('ape'/de OR 'bat'/exp OR 'carnivora'/exp OR 'catarrhini'/de OR 'cercopithecidae''/exp OR ''cetacea''/exp OR 'chimpanzee'/exp OR 'chordata'/de OR 'elephant'/exp OR 'gorilla'/exp OR 'haplorhini'/de OR 'hominid'/de OR 'hylobatidae'/exp OR 'hyrax'/exp OR 'lagomorph'/exp OR 'mammal'/de OR 'marsupial'/exp OR 'monotremate'/exp OR 'orangutan'/exp OR 'placental mammals'/de OR 'platyrrhini'/exp OR 'primate'/de OR 'prosimian'/exp OR 'rodent'/exp OR 'scandentia'/exp OR 'simian'/de OR 'sirenia'/exp OR 'tarsiiform'/exp OR 'ungulate'/exp OR 'vertebrate'/de OR 'xenarthra'/exp OR animals:ti,ab OR animal:ti,ab OR mice:ti,ab OR mus:ti,ab OR mouse:ti,ab OR murine:ti,ab OR woodmouse:ti,ab OR rats:ti,ab OR rat:ti,ab OR murinae:ti,ab OR muridae:ti,ab OR cottonrat:ti,ab OR cottonrats:ti,ab OR hamster:ti,ab OR hamsters:ti,ab OR cricetinae:ti,ab OR rodentia:ti,ab OR rodent:ti,ab OR rodents:ti,ab OR pigs:ti,ab OR pig:ti,ab OR swine:ti,ab OR swines:ti,ab OR piglets:ti,ab OR piglet:ti,ab OR boar:ti,ab OR boars:ti,ab OR "sus scrofa":ti,ab OR ferrets:ti,ab OR ferret:ti,ab OR polecat:ti,ab OR polecats:ti,ab OR "mustela putorius":ti,ab OR "guinea pigs":ti,ab OR "guinea pig":ti,ab OR cavia:ti,ab OR callithrix:ti,ab OR marmo-

set:ti,ab OR marmosets:ti,ab OR cebuella:ti,ab OR hapale:ti,ab OR octodon:ti,ab OR chinchilla:ti,ab OR chinchillas:ti,ab OR gerbillinae:ti,ab OR gerbil:ti,ab OR gerbils:ti,ab OR jird:ti,ab OR jirds:ti,ab OR merione:ti,ab OR meriones:ti,ab OR rabbits:ti,ab OR rabbit:ti,ab OR hares:ti,ab OR hare:ti,ab OR cats:ti,ab OR cat:ti,ab OR felis:ti,ab OR dogs:ti,ab OR dog:ti,ab OR canine:ti,ab OR canines:ti,ab OR canis:ti,ab OR sheep:ti,ab OR sheeps:ti,ab OR mouflon:ti,ab OR mouflons:ti,ab OR ovis:ti,ab OR goats:ti,ab OR goat:ti,ab OR capra:ti,ab OR capras:ti,ab OR rupicapra:ti,ab OR chamois:ti,ab OR haplorhini:ti,ab OR monkey:ti,ab OR monkeys:ti,ab OR anthropoidea:ti,ab OR anthropoids:ti,ab OR saguinus:ti,ab OR tamarin:ti,ab OR tamarins:ti,ab OR leontopithecus:ti,ab OR hominidae:ti,ab OR ape:ti,ab OR apes:ti,ab OR pan:ti,ab OR paniscus:ti,ab OR "pan paniscus":ti,ab OR bonobo:ti,ab OR bonobos:ti,ab OR "pan troglodytes":ti,ab OR gibbon:ti,ab OR gibbons:ti,ab OR siamang:ti,ab OR siamangs:ti,ab OR nomascus:ti,ab OR symphalangus:ti,ab OR chimpanzee:ti,ab OR chimpanzees:ti,ab OR prosimians:ti,ab OR "bush baby":ti,ab OR prosimian:ti,ab OR bush babies:ti,ab OR galagos:ti,ab OR galago:ti,ab OR pongidae:ti,ab OR gorilla:ti,ab OR gorillas:ti,ab OR pongo:ti,ab OR "pongo pygmaeus":ti,ab OR orangutans:ti,ab OR lemur:ti,ab OR lemurs:ti,ab OR lemuridae:ti,ab OR horse:ti,ab OR horses:ti,ab OR pongo:ti,ab OR equus:ti,ab OR cow:ti,ab OR calf:ti,ab OR bull:ti,ab OR chicken:ti,ab OR chickens:ti,ab OR squirrel:ti,ab OR squirrels:ti,ab OR chipmunk:ti,ab OR chipmunks:ti,ab OR suslik:ti,ab OR susliks:ti,ab OR vole:ti,ab OR voles:ti,ab OR lemming:ti,ab OR lemmings:ti,ab OR muskrat:ti,ab OR muskrats:ti,ab OR lemmus:ti,ab OR otter:ti,ab OR otters:ti,ab OR marten:ti,ab OR martens:ti,ab OR martes:ti,ab OR weasel:ti,ab OR badger:ti,ab OR badgers:ti,ab OR ermine:ti,ab OR mink:ti,ab OR minks:ti,ab OR sable:ti,ab OR sables:ti,ab OR gulo:ti,ab OR gulos:ti,ab OR wolverine:ti,ab OR wolverines:ti,ab OR minks:ti,ab OR mustela:ti,ab OR llama:ti,ab OR llamas:ti,ab OR alpaca:ti,ab OR alpacas:ti,ab OR camelid:ti,ab OR camelids:ti,ab OR guanaco:ti,ab OR guanacos:ti,ab OR chiroptera:ti,ab OR chiropteras:ti,ab OR bat:ti,ab OR bats:ti,ab OR fox:ti,ab OR foxes:ti,ab OR donkey:ti,ab OR donkeys:ti,ab OR mule:ti,ab OR mules:ti,ab OR zebra:ti,ab OR zebras:ti,ab OR shrew:ti,ab OR shrews:ti,ab OR bison:ti,ab OR bisons:ti,ab OR buffalo:ti,ab OR buffaloes:ti,ab OR deer:ti,ab OR deers:ti,ab OR bear:ti,ab OR bears:ti,ab OR panda:ti,ab OR pandas:ti,ab OR "wild hog":ti,ab OR "wild boar":ti,ab OR fitchew:ti,ab OR fitch:ti,ab OR beaver:ti,ab OR beavers:ti,ab OR jerboa:ti,ab OR jerboas:ti,ab OR capybara:ti,ab OR capybaras:ti,ab)

Toxline

("117-81-7" OR "117-84-0" OR "131-11-3" OR "131-18-0" OR "27554-26-3" OR "28553-12-0" OR "84-66-2" OR "84-69-5" OR "84-74-2" OR "85-68-7" OR "Avolin" OR "BBPHT" OR "BBzP" OR "bis 2 ethylhexylphthalate" OR "Bis 2-ethylhexyl phthalate" OR "butylbenzylphthalate" OR "butyl-benzylphthalate" OR "BzBP" OR "Dbp" OR "DEP" OR "di n octylphthalate" OR "DiBP" OR "diethylphthalate" OR "di-isononylphthalate" OR "Diisooctylphthalate" OR "Dimethylphthalate" OR "DINP" OR "dioctylphthalate" OR "dipentyl benzene-1,2-dicarboxylate" OR "Dmp" OR "dmp30" OR "DNOP" OR "ENJ 2065" OR "fermine" OR "mipax" OR "mugia" OR "octoil" OR "o-phthalate" OR "o-phthalates" OR "palatinol" OR "sketofax") AND ("Exposure" OR "Exposed" OR "exposures" OR "exposing") AND ("Anogenital" OR "AGD" OR "AGI" OR "ASD" OR "APD" OR "Urogenital" OR "Penile" OR "penis" OR "Anoscrotal" OR "Anopenile" OR "anorectal" OR "Testosterone" OR "androgen" OR "androgens" OR "Hypospadias" OR "hypospadia" OR "Testis" OR "testes" OR (("Anorectal" OR "genital" OR "genitals" OR "testes" OR "rectum") AND ("malformation" OR "malformations" OR "development" OR "abnormalities" OR "abnormality" OR "dysplasia")) OR ("Male" and ("genital" OR "genitals" OR "genitalia")) OR ("Anus" AND ("genital" OR "genitals" OR "genitalia"))) AND (animals OR animal OR mice OR mus OR mouse OR murine OR woodmouse OR rats OR rat OR murinae OR muridae OR cottonrat OR cottonrats OR hamster OR hamsters OR cricetinae OR rodentia OR rodent OR rodents OR pigs OR pig OR swine OR swines OR piglets OR piglet OR boar OR boars OR "sus scrofa" OR ferrets OR ferret OR polecat OR polecats OR "mustela putorius" OR "guinea pigs" OR "guinea pig" OR cavia OR callithrix OR marmoset OR marmosets OR cebuella OR hapale OR octodon OR chinchilla OR chinchillas OR gerbillinae OR gerbil OR gerbils OR jird OR jirds OR merione OR meriones OR rabbits OR rabbit OR hares OR hare OR cats OR cat OR felis OR dogs OR dog OR canine OR canines OR

canis OR sheep OR sheeps OR mouflon OR mouflons OR ovis OR goats OR goat OR capra OR capras OR rupicapra OR chamois OR haplorhini OR monkey OR monkeys OR anthropoidea OR anthropoids OR saguinus OR tamarin OR tamarins OR leontopithecus OR hominidae OR ape OR apes OR pan OR paniscus OR "pan paniscus" OR bonobo OR bonobos OR "pan troglodytes" OR gibbon OR gibbons OR siamang OR siamangs OR nomascus OR symphalangus OR chimpanzee OR chimpanzees OR prosimians OR "bush baby" OR prosimian OR bush babies OR galagos OR galago OR pongidae OR gorilla OR gorillas OR pongo OR "pongo pygmaeus" OR orangutans OR lemur OR lemurs OR lemuridae OR horse OR horses OR pongo OR equus OR cow OR calf OR bull OR chicken OR chickens OR squirrel OR squirrels OR chipmunk OR chipmunks OR suslik OR susliks OR vole OR voles OR lemming OR lemmings OR muskrat OR muskrats OR lemmus OR otter OR otters OR marten OR martens OR martes OR weasel OR badger OR badgers OR ermine OR mink OR minks OR sable OR sables OR gulo OR gulos OR wolverine OR wolverines OR minks OR mustela OR llama OR llamas OR alpaca OR alpacas OR camelid OR camelids OR guanaco OR guanacos OR chiroptera OR chiropteras OR bat OR bats OR fox OR foxes OR donkey OR donkeys OR mule OR mules OR zebra OR zebras OR shrew OR shrews OR bison OR bisons OR buffalo OR buffaloes OR deer OR deers OR bear OR bears OR panda OR pandas OR "wild hog" OR "wild boar" OR fitchew OR fitch OR beaver OR beavers OR jerboa OR jerboas OR capybara OR capybaras)

SECTION C-1c

SCREENING FORMS

Title and Abstract Screening Form

INSTRUCTIONS: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include 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)	

Full-Text Screening Form

INSTRUCTIONS: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include nonhuman mammals	
Study does not report phthalate exposure to one or more of the phthalates listed in the PECO statement	
Study does not report oral exposure to phthalates	
Study does not quantify exposure to phthalates	
Study does not include in utero exposure	
Study does not assess or report anogenital distance, anogenital distance, anoscrotal distance, anopenile distance, hypospadias, or fetal testosterone concentration	
No comparator group (different doses or vehicle-only treatment)	
Not in English	
Other (explanation required)	

SECTION C-1d

DATA EXTRACTION ELEMENTS FOR ANIMAL STUDIES

Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or life stage at start of dosing and at health outcome assessment
	Definition of gestation age for the day after mating (e.g., GD 0 vs GD 1)
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic [e.g., 90 days in a rodent], chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP, or another guideline for study design, conducted under good laboratory practice [GLP] guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls—was expected response observed? (*information bias)
	End point health category (e.g., reproductive)
	End point (e.g., infertility)
	Diagnostic or method to measure end point (*information bias)
	Statistical methods (*information bias)
Results	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. When possible, measures of effect will be converted to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).

	No-observed-effect level (NOEL), lowest-observed-effect level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. Note: The NOEL and LOEL 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 NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias.

SECTION C-1e

RISK OF BIAS QUESTIONS FOR ANIMAL STUDIES

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
• Direct evidence that animals were allocated to any study group including controls using a method with a random component, • **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups. • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green, 2011). • **Note:** Restricted randomization (e.g., blocked randomization) to ensure that particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
• Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that animals were allocated to study groups using a method with a nonrandom component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how animals were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that animals were allocated to study groups using a nonrandom method, including judgment of the investigator, the results of a laboratory test, or a series of tests, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
• Direct evidence that the same vehicle was used in control and experimental animals, • **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that the same vehicle was used in control and experimental animals, • **OR** it is deemed that the vehicle used would not appreciably bias results, • **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the vehicle differed between control and experimental animals, • **OR** authors did not report the vehicle used (record "NR" as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
• Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than were experimental animals, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of the study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study. • **Note:** Acceptable handling of attrition includes very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or 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; missing outcomes is not enough to impact the effect estimate. • **OR** missing data have been imputed using appropriate methods (ensuring that characteristics of animals are not significantly different from animals retained in the analysis).
Probably Low Risk of Bias (+)
• Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, • **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of animals was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of animals was unacceptably large and not adequately addressed. • **Note:** Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that the exposure to the phthalate was independently characterized (including purity, stability, and compliance with the treatment, if applicable) and confirmed generally as ≥98% purity, • **OR** direct evidence that all individual congeners were independently assessed for purity if a "mixture" is developed by the researchers, • **OR** the mixture should be independently assessed and non-target congeners or other impurities confirmed to contribute less than 2% (purity is ≥98%), • **AND** that exposure was consistently administered (i.e., with the same method and time frame) across treatment groups, • **AND** for gavage, dietary, or drinking water studies, that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups, • **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure to the phthalate was independently characterized (including purity, stability, and compliance with the treatment, if applicable) and confirmed generally as ≥98% (i.e., the supplier of the chemical provides documentation of the purity of the chemical), • **OR** indirect evidence that all individual congeners were independently assessed for purity if a "mixture" is developed by the researchers (the supplier of the chemical provides documentation of the purity of each chemical) and non-target congeners/impurities confirmed as less than 98%, • **OR** the mixture is provided by a supplier and the supplier provides documentation of the purity of the mixture with non-target congeners/impurities confirmed to contribute less than 2% (purity is ≥98%), • **OR** direct evidence that the purity of the congener(s) was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results, • **AND** that exposure was consistently administered (i.e., with the same method and time frame) across treatment groups, • **AND** for dietary or drinking water studies, no information is provided on consumption or internal dose metrics, • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record "NR" as basis for answer),

• **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
• Direct evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., commercial RIA or ELISA kit for fetal testosterone; micrometer caliper or reticule micrometer for AGD), • **AND** assessed at the same length of time (i.e., same day of life) after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **Note:** Fetal testosterone measured in testes or media in which testes had been incubated.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable), • **AND** assessed at the same length of time (i.e., same day of life) after initial exposure in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, 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, which is more likely to apply to objective outcome measures. • **Note:** Fetal testosterone measured in testes, media in which testes had been incubated, or fetal blood-derived media.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer), • **OR** in results or analyses of AGD the measurement method for AGD not reported.
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes 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 in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes 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 a composite score without individual outcome components or outcomes reported using measurements, analysis methods, or subsets of the data (e.g., subscales) that were not prespecified or reporting outcomes not prespecified, or that unplanned analyses were included that would appreciably bias results.

11. Was litter or litter effects considered appropriately in the statistical analyses and were there no other potential threats to internal validity?

Because this evaluation is focused on developmental exposure, this question was added to address litter effects in data analysis. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk of bias considerations that do not fit under the other questions.

Definitely Low Risk of Bias (++)
• Direct evidence that litter effects were appropriately considered in the study design or analysis, using one of the following approaches: • The dam used as the statistical unit of analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects were appropriately considered in the analysis AND the statistical method was stated.
Probably Low Risk of Bias (+)
• Indirect evidence that litter effects were appropriately considered in the study design or analysis, using one of the following approaches: • The dam used as the statistical unit of analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects were appropriately considered in the analysis BUT the statistic method used to address litter effects was not stated.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that litter effects were not appropriately considered in the study design or analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects **were not** considered in the statistical analysis.
Definitely High Risk of Bias (--)
• Direct evidence that litter effects were not appropriately considered in the study design or analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects **were not** considered in the statistical analysis.

SECTION C-1f

AMENDMENTS TO THE PROTOCOL

Addition of an Exclusion Criteron for Full-Text Screening

The following criterion for excluding studies at the full-text screening level was added on September 15, 2016, after the protocol was peer reviewed:

Study involved rodents exposed to a single high dose (≥500 mg/kg/day).

The committee judged that a study testing only a single high dose level would not be useful for a systematic review intended to address the low-dose toxicity of phthalates.

Additions to the Review Team

The following committee members were added to the review team to supplement expertise:

- **Weihsueh Chiu** is a professor in the Department of Veterinary Integrative Biosciences in the College of Veterinary Medicine and Biomedical Sciences at Texas A&M University. Before joining the university, he worked at the US Environmental Protection Agency (EPA) for more than 14 years, most recently as chief of the Toxicity Pathways Branch in the Integrated Risk Information System (IRIS) Division of the National Center for Environmental Assessment. His research has focused on human health risk assessment, particularly with respect to toxicokinetics, mechanisms of toxicity, physiologically based pharmacokinetic modeling, dose-response assessment, and characterizing uncertainty and variability. He led the development of EPA's 2011 IRIS assessment of trichloroethylene, which pioneered the use of probabilistic methods for characterizing uncertainty and variability in toxicokinetics and dose-response. He is currently a member of the NRC's Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures. Dr. Chiu received a PhD in physics from Princeton University.
- **Katrina Waters** is deputy director of the Biological Sciences Division at the Pacific Northwest National Laboratory. Her research interests are focused on the integration of genomics, proteomics, metabolomics and high-throughput screening data to enable predictive mechanistic modeling of disease and toxicity pathways. She currently serves on EPA's Board of Scientific Counselors Subcommittee on Chemical Safety for Sustainability and the US Food and Drug Administration's Scientific Advisory Board to the National Center for Toxicological Research. She recently served on the NRC's Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures. Dr. Waters received a PhD in biochemistry from the University of Wisconsin–Madison, and did a postdoctoral fellowship on endocrine disruptors at the Chemical Industry Institute of Toxicology.

SECTION C-2

Results of Literature Searches for Animal Studies on the Effects of Phthalates on Male Reproductive-Tract Development

Literature searches were performed on August 15, 2016, using the search strategy presented in the Phthalate (Animal) Systematic Review Protocol (Section C-1). A summary of the results is presented below.

Embase:	754
PubMed:	521
Toxline:	865

Total citations found:	2,140
Duplicates removed:	613
Total unique citations:	1,527

SECTION C-3

Funding Sources of the Animal Studies on Phthalates and Male Reproductive-Tract Development

Sources of funding were used to evaluate publication bias in terms of whether a particular sector funded more studies than another.

Reference	Government	Industry	Other	Unknown
Adamsson et al. 2009	X (European Commission)		X (Academy of Finland, Sigrid Juselius Foundation, The Finnish Concordia Fund, Turku University Hospital)	
Ahmad et al. 2014	X (ICMR-India)			
Andrade et al. 2006	X (Germany)			
Ashby et al. 1997		X (Zeneca, UK)		
Aso et al. 2005	X (Japan)			
Barlow et al. 2004	X (NIEHS)	X (American Chemistry Council)		
Beverly et al. 2014	X (EPA, NIEHS)			
Boberg et al. 2011	X (Nordic Council of Ministers, EU)			
Borch et al. 2004	X (Denmark)			
Borch et al. 2006	X (Nordic Council of Ministers, EU)			
Christiansen et al. 2009	X (Danish EPA, EU)			
Christiansen et al. 2010	X (EU)			
Clewell et al. 2009	X (NIEHS)	X (American Chemistry Council)		
Clewell et al. 2013		X (ExxonMobil Biochemical Sciences)		
Culty et al. 2008	X (NIEHS)			
Do et al. 2012				X
Drake et al. 2009	X (UK Medical Research Council, EU)			
Ema and Miyawaki 2002	X (Japan)			
Ema et al. 1998				X
Ema et al. 2000	X (Japan)			
Fujii et al. 2005	X (Japan)			
Furr et al. 2014	X (NIEHS, NTP, NIH)			
Giribabu et al. 2014	X (ICMR-India)			
Gray et al. 2009	X (NSF)			
Hannas et al. 2011a	X (EPA)			
Hannas et al. 2011b	X (EPA)			
Hannas et al. 2012	X (EPA, NTP, NAS)			
Howdeshell et al. 2008	X (EPA)			
Jarfelt et al. 2005	X (Denmark)			
Jiang et al. 2007	X (China)			
Johnson et al. 2007		X (American Chemistry Council)		
Johnson et al. 2011	X (NIH)			

Jones et al. 2015	X (Canada)			
Kim et al. 2010	X (Korea)			
Klinefelter et al. 2012				This research didn't receive any funding.
Kuhl et al. 2007	X (NIH)			
Lee et al. 2004	X (Japan)			
Lehmann et al. 2004	X (NIH)			
Li et al. 2009	X (China)			
Li et al. 2013	X (China)			
L. Li et al. 2015	X (China)			
N. Li et al. 2015				There are no funding sources to declare.
Lin et al. 2008	X (NIEHS)			
Lin et al. 2009	X (NIEHS)			
Liu et al. 2008	X (China)			
MacLeod et al. 2010	X (UK Medical Research Council, EU)			
Mahood et al. 2007	X (EU)			
Martino-Andrade et al. 2009	X (Brazil)			
Masutomi et al. 2003	X (Japan)			
McKinnel et al. 2009	X (UK Medical Research Council)			
Moore et al. 2001	X (NIH)			
Mylchreest et al. 1998		Chemical Industry Institute of Toxicology		
Mylchreest et al. 1999		Chemical Industry Institute of Toxicology		
Mylchreest et al. 2000		Chemical Industry Institute of Toxicology		
Nagao et al. 2000	X (Japan)			
Pocar et al. 2012	X (Italy, EU)			
Saillenfait et al. 2008	X (INRS-France)			
Saillenfait et al. 2009	X (INRS-France)			
Saillenfait et al. 2011	X (INRS-France)			
Saillenfait et al. 2013a	X (INRS-France)			
Saillenfait et al. 2013b	X (INSERM, INRS-France)			
Scarano et al. 2010			X (State of Sao Paulo Research Foundation)	
Struve et al. 2009		X (American Chemistry Council)		
Tyl et al. 2004		X (European Council for Plasticisers and Intermediates)		
van den Driesche et al. 2012	X (UK Medical Research Council, EU)			
Vo et al. 2009	X (Korea)			
Wolfe and Layton 2005	X (NTP/NIEHS)			
Wolfe and Patel 2002	X (DHHS)			
Zhang et al. 2004	X (China)			
Zhang et al. 2013				X

SECTION C-4

Confidence Ratings for the Body of Evidence from Animal Studies of Phthalates and Anogenital Distance (AGD), Fetal Testosterone, and Hypospadias

The confidence in the body of evidence from animal studies on phthalates and male reproductive-tract development was rated in accordance with the OHAT Guidance (NTP 2015) specified in Section C-1. The results for di(2-ethylhexyl) phthalate/diethylhexyl phthalate (DEHP) are presented first, and the remaining phthalates are subsequently presented in alphabetical order.

DEHP and AGD

Nineteen animal studies of DEHP and AGD were available; 3 used the mouse model and 16 used the rat model.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. Two of the three mouse studies did not control or account for litter effects, and the studies had issues with outcome assessment and lack of blinding of the researchers to the study groups during outcome assessment (see Figure C4-1). Six of 16 rat studies did not account for litter effects, and most of the studies also had issues with outcome assessment and blinding of the researchers (see Figure C4-2).

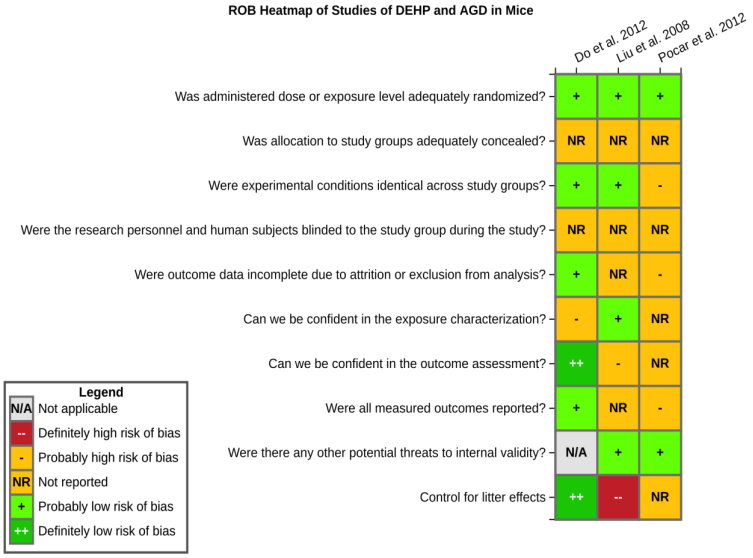


FIGURE C4-1 Risk of bias heatmap of studies of DEHP and AGD in mice. In HAWC: https://hawcproject.org/summary/visual/302/.

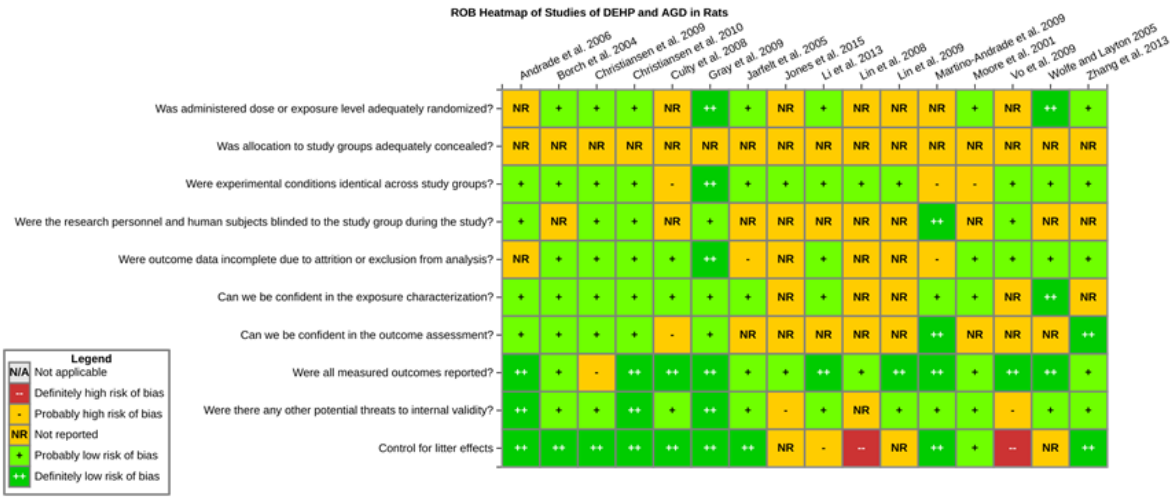


FIGURE C4-2 Risk of bias heatmap of studies of DEHP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/319/.

- **Unexplained inconsistencies:** No downgrade. Although there appeared to be heterogeneity in the results (see Figure C4-3), most of it could be explained by dose, species, or strain differences. Meta-analyses of the data found no important heterogeneity in the rat or the mouse data (see Appendix C, Section C-5), further supporting the decision not to downgrade.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. Mean versus standard deviation for most studies reflects reasonable precision (see Figure C4-3). Meta-analyses of the data found a statistically significant summary overall estimate for rats but not for mice (see Appendix C, Section C-5). Because the mouse studies account for a small percentage of the overall body of evidence (3 of 19 studies), confidence was not downgraded for imprecision.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Meta-analysis of the data showed that, in rats, the effects could be considered large and robust, with overall summary estimates having z-scores of ≥7.0 (see Appendix C, Section C-5). The effect sizes were robust to multiple sensitivity analyses.
- **Dose-response:** Upgraded. Although the visualization in Figure C4-4 suggests some inconsistency in dose response across studies, an upgrade is supported by the meta-analysis of the rat data, which found statistically significant linear trends in $\log_{10}$ (dose) or dose. The results were robust to multiple sensitivity analyses.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade.

DEHP and Fetal Testosterone

Twelve animal studies of DEHP and fetal testosterone were available; 11 used the rat model and 1 used the mouse model.

FIGURE C4-3 Data pivot of animal studies of DEHP and AGD sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/z1-phthalate-effect-agd-all/.

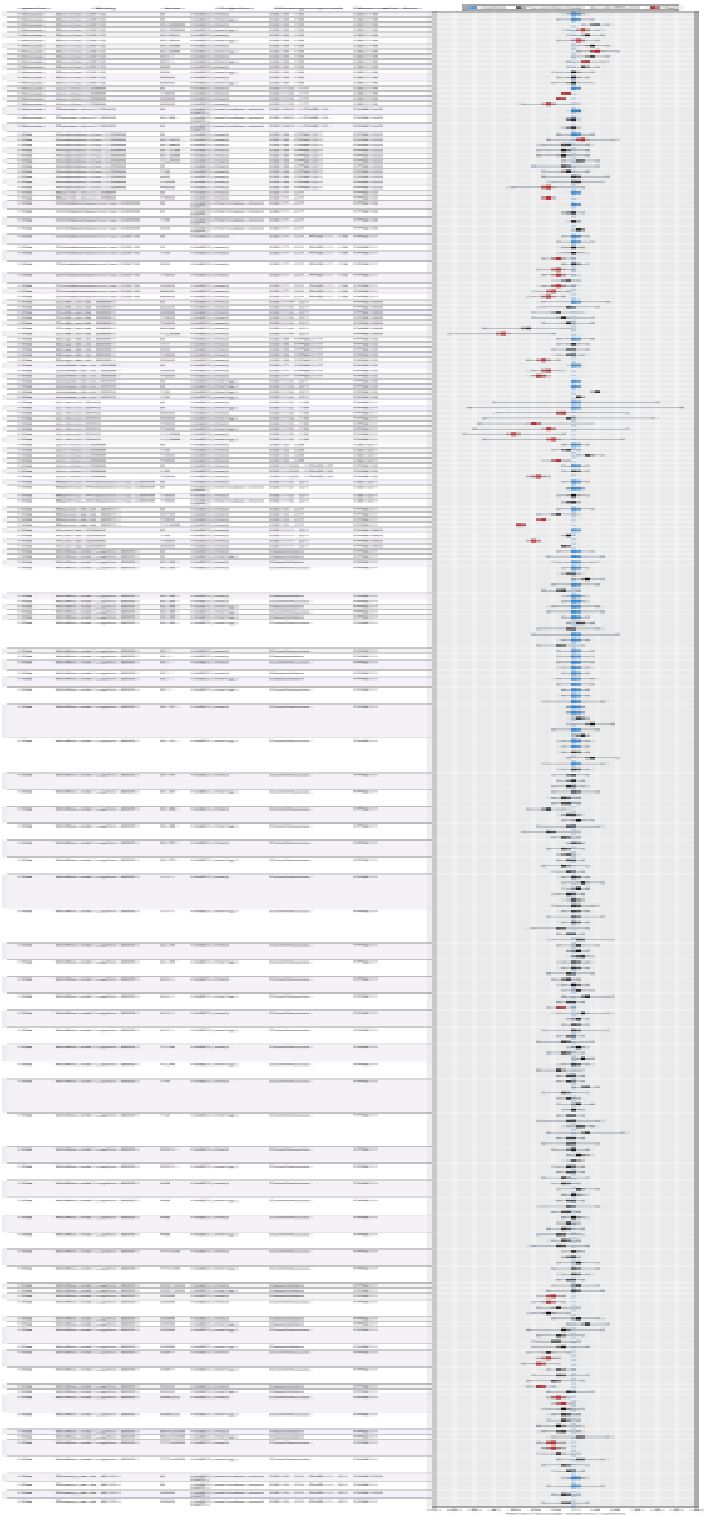


FIGURE C4-4 Data pivot of animal studies of DEHP and AGD sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/z1-dehp-effect-agd-dose-response-new/.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. Most studies accounted for litter effects and used reliable methods of measuring fetal testosterone. See Figure C4-5.
- **Unexplained inconsistencies:** No downgrade. Consistent decrease in fetal testosterone across studies, with a few exceptions that can be explained by study design features (e.g., examining testosterone in fetal plasma, which might have technical difficulties). See Figure C4-6. A meta-analysis of the data also supported the decision not to downgrade (see Appendix C, Section C-5).
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. A meta-analysis of the data found a statistically significant summary overall estimate, linear trend in $\log_{10}$(dose), and linear trend in dose, which were robust to multiple sensitivity analyses.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. High dose groups reflect a relatively large magnitude of change (about 75-90%) across several studies. See Figure C4-6. A meta-analysis of the data also supported the decision to upgrade (see Appendix C, Section C-5).
- **Dose-response:** Upgraded. Several studies reflect a dose response in the same dose ranges (see Figure C4-7). A meta-analysis of the data also supported the decision to upgrade (see Appendix C, Section C-5).
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only one mouse study was available so cross-species consistency could not be evaluated. Results are generally consistent across studies in the high dose range.

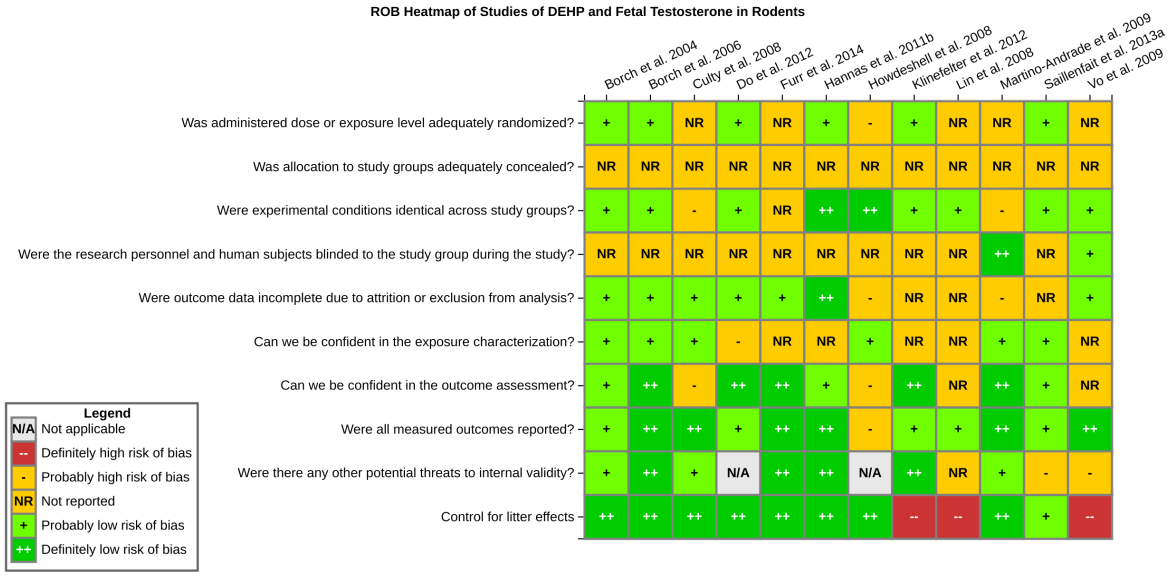


FIGURE C4-5 Risk of bias heatmap of studies of DEHP and fetal testosterone in rodents. In HAWC: https://hawcproject.org/summary/visual/362/.

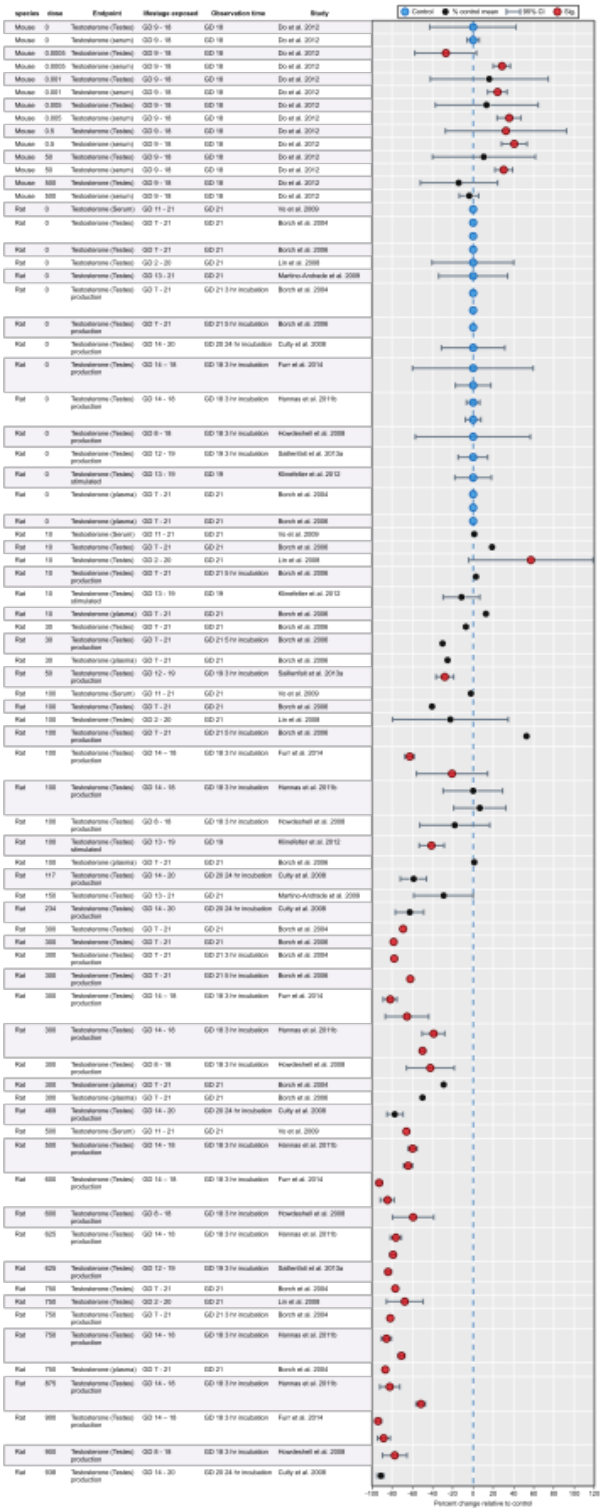


FIGURE C4-6 Data pivot of animal studies of DEHP and fetal testosterone sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-agd/.

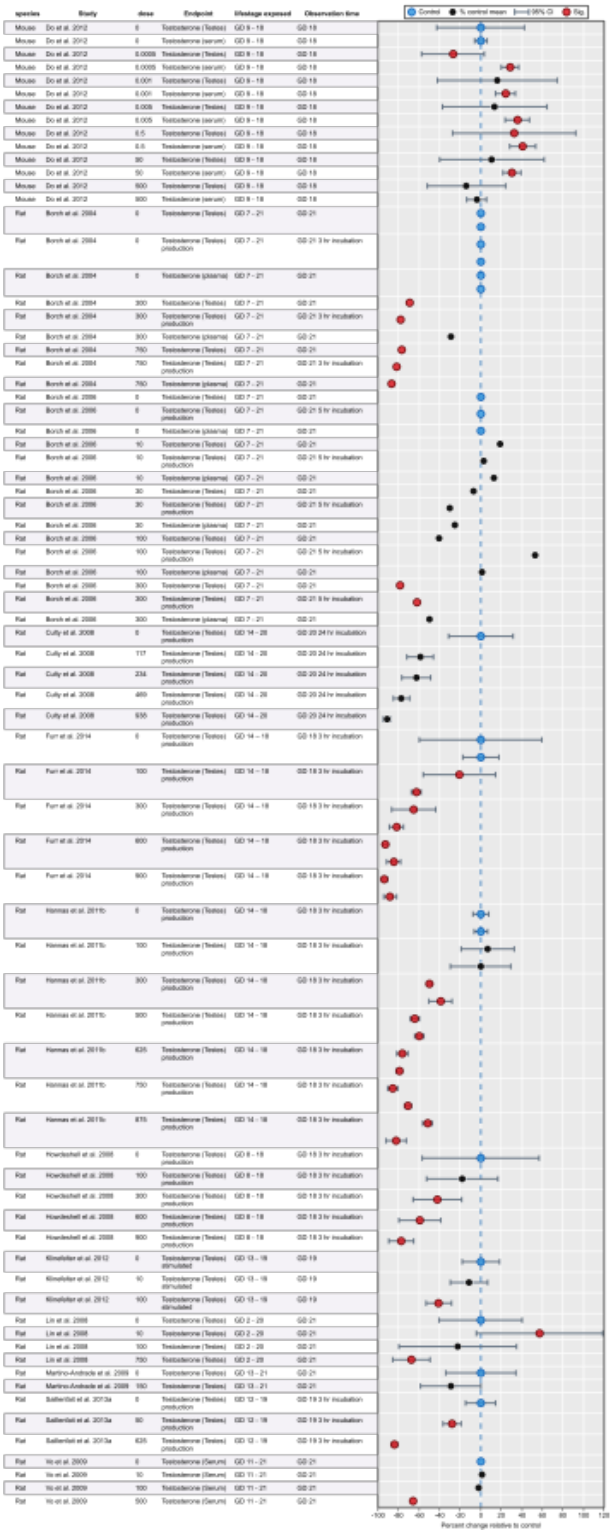


FIGURE C4-7 Data pivot of animal studies of DEHP and fetal testosterone sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-testosterone-dose-response/.

DEHP and Hypospadias

Nine animal studies of DEHP and hypospadias were available; 8 used the rat model and 1 used the mouse model.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. Over half of the studies had a probably high risk of bias rating because they lacked reporting on the outcome assessment. Other concerns were related to whether the researchers were blinded to the study groups during outcome assessment and not controlling for litter effects. See Figure C4-8.
- **Unexplained inconsistencies:** Downgraded. Incidence of hypospadias is not consistent across studies within similar dose ranges (e.g., Christiansen et al. [2009, 2010] and Jarfelt et al. [2005] show no increased incidence at doses of 750 mg/kg-day or higher). See Figure C4-9.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. No confidence intervals for incidence data, but no hypospadias in control groups. See Figure C4-9.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. Incidence of hypospadias is not consistently large in magnitude across studies for high dose groups. See Figure C4-9.
- **Dose-response:** No upgrade. Dose response noted for a couple of the studies but not consistently across studies. See Figure C4-10.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgraded. Only one mouse study was available so cross-species consistency could not be evaluated. Results are generally not consistent across studies.
- **Rare outcome:** Upgraded. Background control incidence of hypospadias was reported as zero across all studies, so any positive finding was considered treatment related.

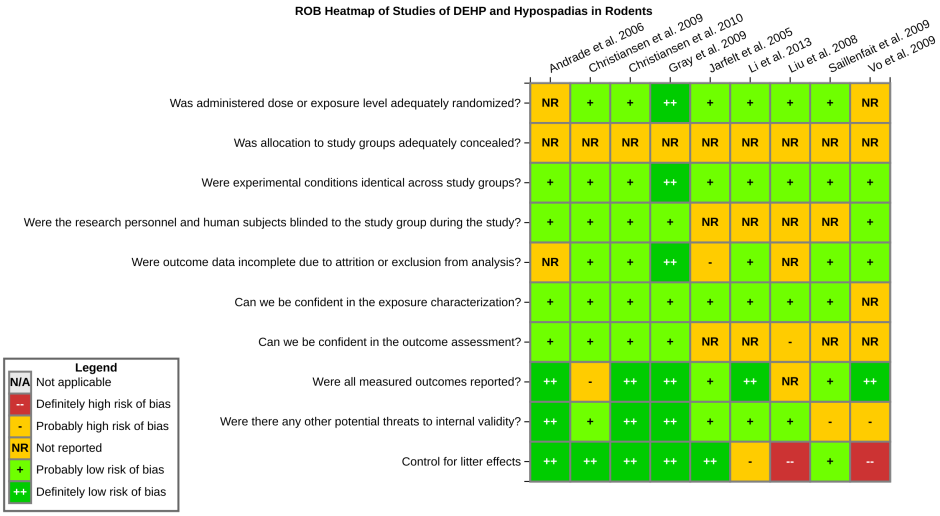
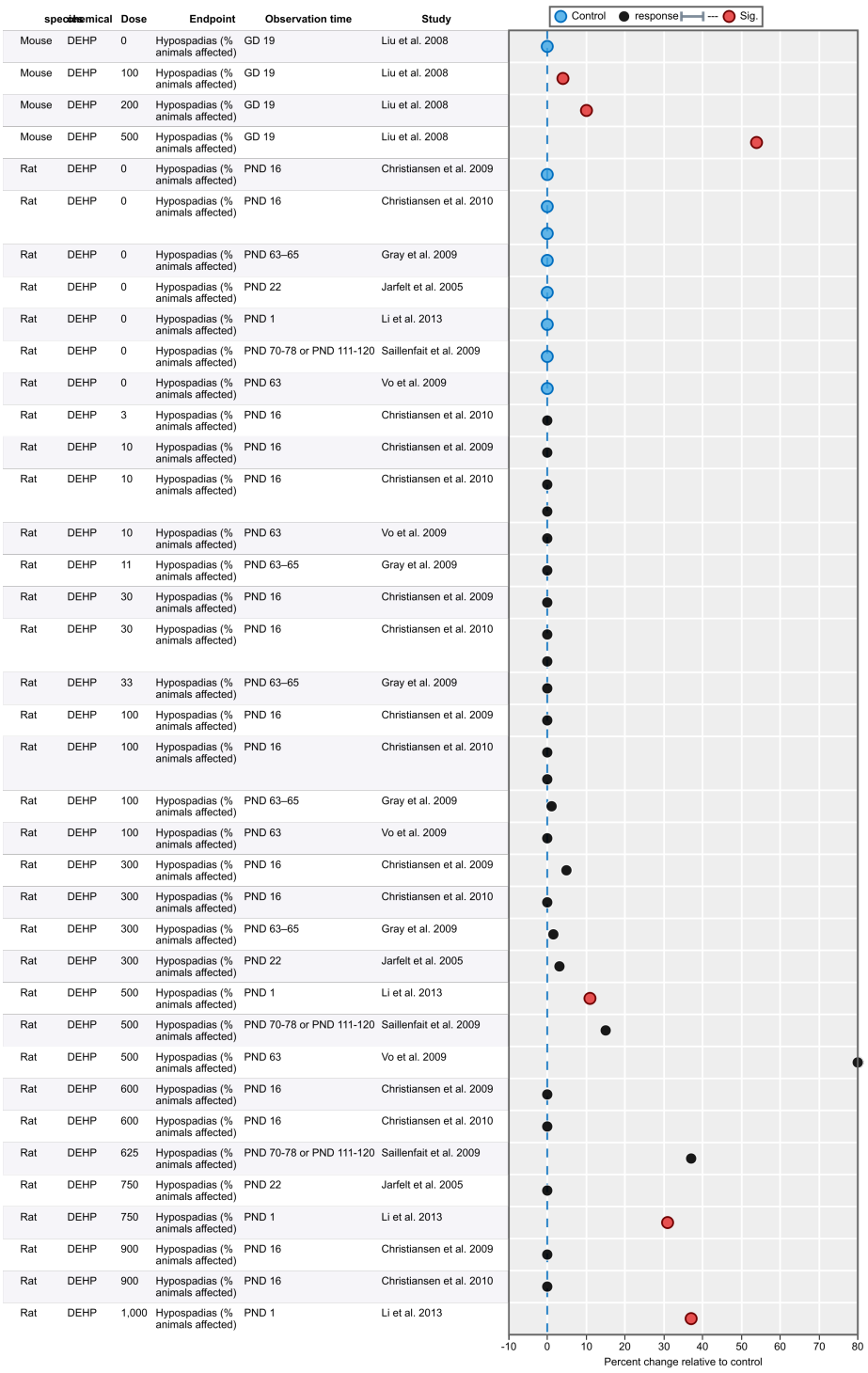


FIGURE C4-8 Risk of bias heatmap of studies of DEHP and hypospadias in rodents. In HAWC: https://hawcproject.org/summary/visual/360/.



species	chemical	Dose	Endpoint	Observation time	Study
Mouse	DEHP	0	Hypospadias (% animals affected)	GD 19	Liu et al. 2008
Mouse	DEHP	100	Hypospadias (% animals affected)	GD 19	Liu et al. 2008
Mouse	DEHP	200	Hypospadias (% animals affected)	GD 19	Liu et al. 2008
Mouse	DEHP	500	Hypospadias (% animals affected)	GD 19	Liu et al. 2008
Rat	DEHP	0	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	0	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	0	Hypospadias (% animals affected)	PND 63–65	Gray et al. 2009
Rat	DEHP	0	Hypospadias (% animals affected)	PND 22	Jarfelt et al. 2005
Rat	DEHP	0	Hypospadias (% animals affected)	PND 1	Li et al. 2013
Rat	DEHP	0	Hypospadias (% animals affected)	PND 70-78 or PND 111-120	Saillenfait et al. 2009
Rat	DEHP	0	Hypospadias (% animals affected)	PND 63	Vo et al. 2009
Rat	DEHP	3	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	10	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	10	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	10	Hypospadias (% animals affected)	PND 63	Vo et al. 2009
Rat	DEHP	11	Hypospadias (% animals affected)	PND 63–65	Gray et al. 2009
Rat	DEHP	30	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	30	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	33	Hypospadias (% animals affected)	PND 63–65	Gray et al. 2009
Rat	DEHP	100	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	100	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	100	Hypospadias (% animals affected)	PND 63–65	Gray et al. 2009
Rat	DEHP	100	Hypospadias (% animals affected)	PND 63	Vo et al. 2009
Rat	DEHP	300	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	300	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	300	Hypospadias (% animals affected)	PND 63–65	Gray et al. 2009
Rat	DEHP	300	Hypospadias (% animals affected)	PND 22	Jarfelt et al. 2005
Rat	DEHP	500	Hypospadias (% animals affected)	PND 1	Li et al. 2013
Rat	DEHP	500	Hypospadias (% animals affected)	PND 70-78 or PND 111-120	Saillenfait et al. 2009
Rat	DEHP	500	Hypospadias (% animals affected)	PND 63	Vo et al. 2009
Rat	DEHP	600	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	600	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	625	Hypospadias (% animals affected)	PND 70-78 or PND 111-120	Saillenfait et al. 2009
Rat	DEHP	750	Hypospadias (% animals affected)	PND 22	Jarfelt et al. 2005
Rat	DEHP	750	Hypospadias (% animals affected)	PND 1	Li et al. 2013
Rat	DEHP	900	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2009
Rat	DEHP	900	Hypospadias (% animals affected)	PND 16	Christiansen et al. 2010
Rat	DEHP	1,000	Hypospadias (% animals affected)	PND 1	Li et al. 2013

FIGURE C4-9 Data pivot of animal studies of DEHP and hypospadias (% animals affected) sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-animals-affected/.

The following links have additional visualizations presenting data on hypospadias in terms of the percentage of litters affected (https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-litters-affected/) or litter incidence (https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-litter-incidence/).

species	Study	Dose	Endpoint	lifestage exposed	Observation time
Mouse	Liu et al. 2008	0	Hypospadias (% animals affected)	GD 12 - 17	GD 19
Mouse	Liu et al. 2008	100	Hypospadias (% animals affected)	GD 12 - 17	GD 19
Mouse	Liu et al. 2008	200	Hypospadias (% animals affected)	GD 12 - 17	GD 19
Mouse	Liu et al. 2008	500	Hypospadias (% animals affected)	GD 12 - 17	GD 19
Rat	Christiansen et al. 2009	0	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	10	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	30	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	100	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	300	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	600	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2009	900	Hypospadias (% animals affected)	GD 7 - PND 16	PND 16
Rat	Christiansen et al. 2010	0	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	3	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	10	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	30	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	100	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	300	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	600	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Christiansen et al. 2010	900	Hypospadias (% animals affected)	GD 7 - 21, PND 1 - 16	PND 16
Rat	Gray et al. 2009	0	Hypospadias (% animals affected)	GD 8 - PND 17	PND 63–65
Rat	Gray et al. 2009	11	Hypospadias (% animals affected)	GD 8 - PND 17	PND 63–65
Rat	Gray et al. 2009	33	Hypospadias (% animals affected)	GD 8 - PND 17	PND 63–65
Rat	Gray et al. 2009	100	Hypospadias (% animals affected)	GD 8 - PND 17	PND 63–65
Rat	Gray et al. 2009	300	Hypospadias (% animals affected)	GD 8 - PND 17	PND 63–65
Rat	Jarfelt et al. 2005	0	Hypospadias (% animals affected)	GD 7 - PND 17	PND 22
Rat	Jarfelt et al. 2005	300	Hypospadias (% animals affected)	GD 7 - PND 17	PND 22
Rat	Jarfelt et al. 2005	750	Hypospadias (% animals affected)	GD 7 - PND 17	PND 22
Rat	Li et al. 2013	0	Hypospadias (% animals affected)	GD 12 - 19	PND 1
Rat	Li et al. 2013	500	Hypospadias (% animals affected)	GD 12 - 19	PND 1
Rat	Li et al. 2013	750	Hypospadias (% animals affected)	GD 12 - 19	PND 1
Rat	Li et al. 2013	1,000	Hypospadias (% animals affected)	GD 12 - 19	PND 1
Rat	Saillenfait et al. 2009	0	Hypospadias (% animals affected)	GD 12 - 21	PND 70-78 or PND 111-120
Rat	Saillenfait et al. 2009	500	Hypospadias (% animals affected)	GD 12 - 21	PND 70-78 or PND 111-120
Rat	Saillenfait et al. 2009	625	Hypospadias (% animals affected)	GD 12 - 21	PND 70-78 or PND 111-120
Rat	Vo et al. 2009	0	Hypospadias (% animals affected)	GD 11 - 21	PND 63
Rat	Vo et al. 2009	10	Hypospadias (% animals affected)	GD 11 - 21	PND 63
Rat	Vo et al. 2009	100	Hypospadias (% animals affected)	GD 11 - 21	PND 63
Rat	Vo et al. 2009	500	Hypospadias (% animals affected)	GD 11 - 21	PND 63

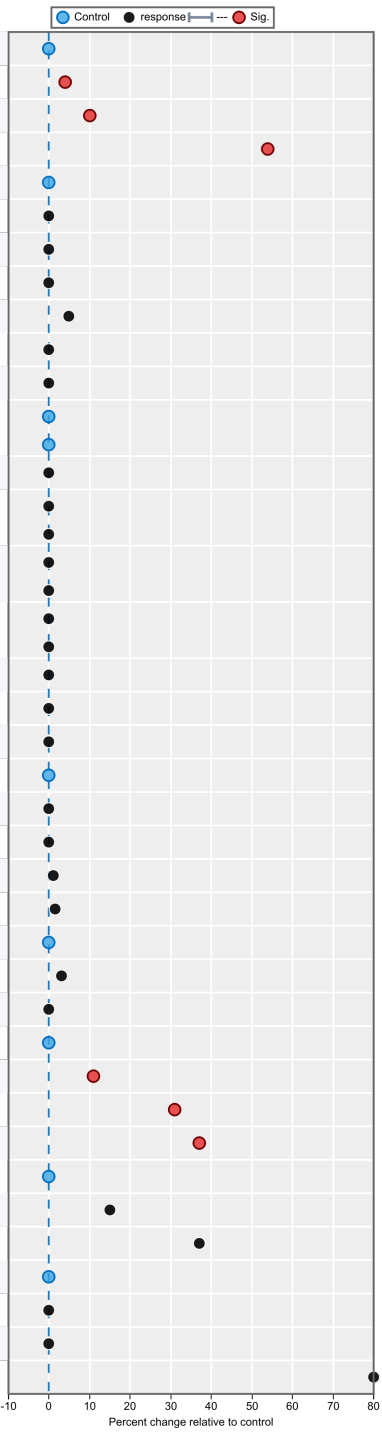


FIGURE C4-10 Data pivot of animal studies of DEHP and hypospadias (% animals affected) sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-animals-affected-dose-resp/.

The following links have additional visualizations presenting data on hypospadias in terms of the percentage of litters affected (https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-litters-affected-dose-resp/) or litter incidence (https://hawcproject.org/summary/data-pivot/assessment/351/dehp-effect-hypospadias-litter-incidence-dose-resp/).

BzBP and AGD

Six studies of BzBP and AGD in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. All the studies had ratings of probably high risk of bias or definitely high risk of bias in at least one of the key issues considered, and all had multiple risk of bias issues. See Figure C4-11.
- **Unexplained inconsistencies:** No downgrade. Consistent dose response across most studies with the exception of the study by Aso et al. (2005), which could be explained by study design features. See Figure C4-12.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. Mean versus standard deviation for the studies reflects reasonable precision. See Figure C4-12.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Three studies reflect relatively large magnitude of change (about 20-40%) in the same dose range. See Figure C4-12.
- **Dose-response:** Upgraded. Most studies reflect a dose response in the same dose range. See Figure C4-13.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

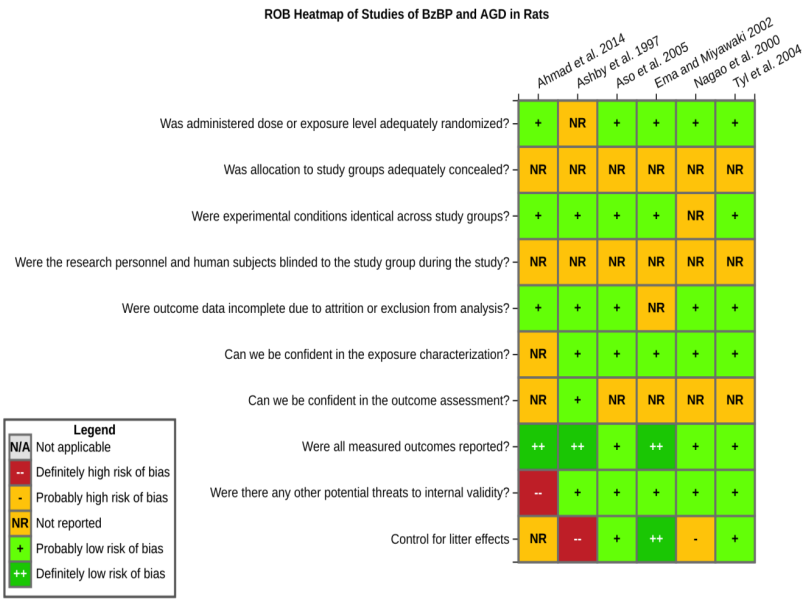
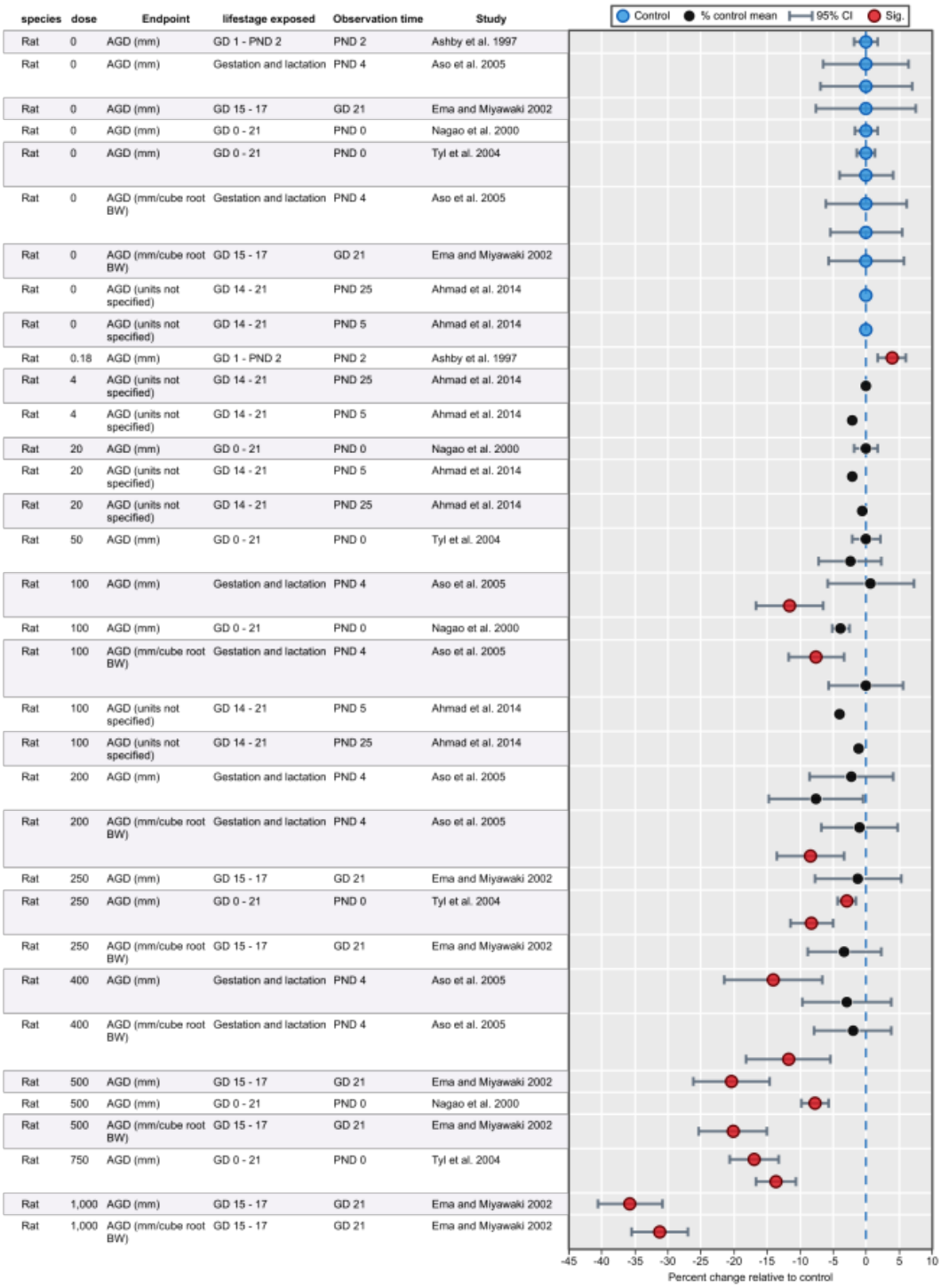


FIGURE C4-11 Risk of bias heatmap of studies of BzBP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/323/.



species	dose	Endpoint	lifestage exposed	Observation time	Study
Rat	0	AGD (mm)	GD 1 - PND 2	PND 2	Ashby et al. 1997
Rat	0	AGD (mm)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	0	AGD (mm)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	0	AGD (mm)	GD 0 - 21	PND 0	Nagao et al. 2000
Rat	0	AGD (mm)	GD 0 - 21	PND 0	Tyl et al. 2004
Rat	0	AGD (mm/cube root BW)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	0	AGD (mm/cube root BW)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	0	AGD (units not specified)	GD 14 - 21	PND 25	Ahmad et al. 2014
Rat	0	AGD (units not specified)	GD 14 - 21	PND 5	Ahmad et al. 2014
Rat	0.18	AGD (mm)	GD 1 - PND 2	PND 2	Ashby et al. 1997
Rat	4	AGD (units not specified)	GD 14 - 21	PND 25	Ahmad et al. 2014
Rat	4	AGD (units not specified)	GD 14 - 21	PND 5	Ahmad et al. 2014
Rat	20	AGD (mm)	GD 0 - 21	PND 0	Nagao et al. 2000
Rat	20	AGD (units not specified)	GD 14 - 21	PND 5	Ahmad et al. 2014
Rat	20	AGD (units not specified)	GD 14 - 21	PND 25	Ahmad et al. 2014
Rat	50	AGD (mm)	GD 0 - 21	PND 0	Tyl et al. 2004
Rat	100	AGD (mm)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	100	AGD (mm)	GD 0 - 21	PND 0	Nagao et al. 2000
Rat	100	AGD (mm/cube root BW)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	100	AGD (units not specified)	GD 14 - 21	PND 5	Ahmad et al. 2014
Rat	100	AGD (units not specified)	GD 14 - 21	PND 25	Ahmad et al. 2014
Rat	200	AGD (mm)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	200	AGD (mm/cube root BW)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	250	AGD (mm)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	250	AGD (mm)	GD 0 - 21	PND 0	Tyl et al. 2004
Rat	250	AGD (mm/cube root BW)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	400	AGD (mm)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	400	AGD (mm/cube root BW)	Gestation and lactation	PND 4	Aso et al. 2005
Rat	500	AGD (mm)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	500	AGD (mm)	GD 0 - 21	PND 0	Nagao et al. 2000
Rat	500	AGD (mm/cube root BW)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	750	AGD (mm)	GD 0 - 21	PND 0	Tyl et al. 2004
Rat	1,000	AGD (mm)	GD 15 - 17	GD 21	Ema and Miyawaki 2002
Rat	1,000	AGD (mm/cube root BW)	GD 15 - 17	GD 21	Ema and Miyawaki 2002

FIGURE C4-12 Data pivot of animal studies of BzBP and AGD in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-agd/.

species	Study	dose	Endpoint	lifestage exposed	Observation time
Rat	Ahmad et al. 2014	0	AGD (units not specified)	GD 14 - 21	PND 5
Rat	Ahmad et al. 2014	0	AGD (units not specified)	GD 14 - 21	PND 25
Rat	Ahmad et al. 2014	4	AGD (units not specified)	GD 14 - 21	PND 5
Rat	Ahmad et al. 2014	4	AGD (units not specified)	GD 14 - 21	PND 25
Rat	Ahmad et al. 2014	20	AGD (units not specified)	GD 14 - 21	PND 5
Rat	Ahmad et al. 2014	20	AGD (units not specified)	GD 14 - 21	PND 25
Rat	Ahmad et al. 2014	100	AGD (units not specified)	GD 14 - 21	PND 5
Rat	Ahmad et al. 2014	100	AGD (units not specified)	GD 14 - 21	PND 25
Rat	Ashby et al. 1997	0	AGD (mm)	GD 1 - PND 2	PND 2
Rat	Ashby et al. 1997	0.18	AGD (mm)	GD 1 - PND 2	PND 2
Rat	Aso et al. 2005	0	AGD (mm)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	0	AGD (mm/cube root BW)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	100	AGD (mm)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	100	AGD (mm/cube root BW)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	200	AGD (mm)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	200	AGD (mm/cube root BW)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	400	AGD (mm)	Gestation and lactation	PND 4
Rat	Aso et al. 2005	400	AGD (mm/cube root BW)	Gestation and lactation	PND 4
Rat	Ema and Miyawaki 2002	0	AGD (mm)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	0	AGD (mm/cube root BW)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	250	AGD (mm)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	250	AGD (mm/cube root BW)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	500	AGD (mm)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	500	AGD (mm/cube root BW)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	1,000	AGD (mm)	GD 15 - 17	GD 21
Rat	Ema and Miyawaki 2002	1,000	AGD (mm/cube root BW)	GD 15 - 17	GD 21
Rat	Nagao et al. 2000	0	AGD (mm)	GD 0 - 21	PND 0
Rat	Nagao et al. 2000	20	AGD (mm)	GD 0 - 21	PND 0
Rat	Nagao et al. 2000	100	AGD (mm)	GD 0 - 21	PND 0
Rat	Nagao et al. 2000	500	AGD (mm)	GD 0 - 21	PND 0
Rat	Tyl et al. 2004	0	AGD (mm)	GD 0 - 21	PND 0
Rat	Tyl et al. 2004	50	AGD (mm)	GD 0 - 21	PND 0
Rat	Tyl et al. 2004	250	AGD (mm)	GD 0 - 21	PND 0
Rat	Tyl et al. 2004	750	AGD (mm)	GD 0 - 21	PND 0

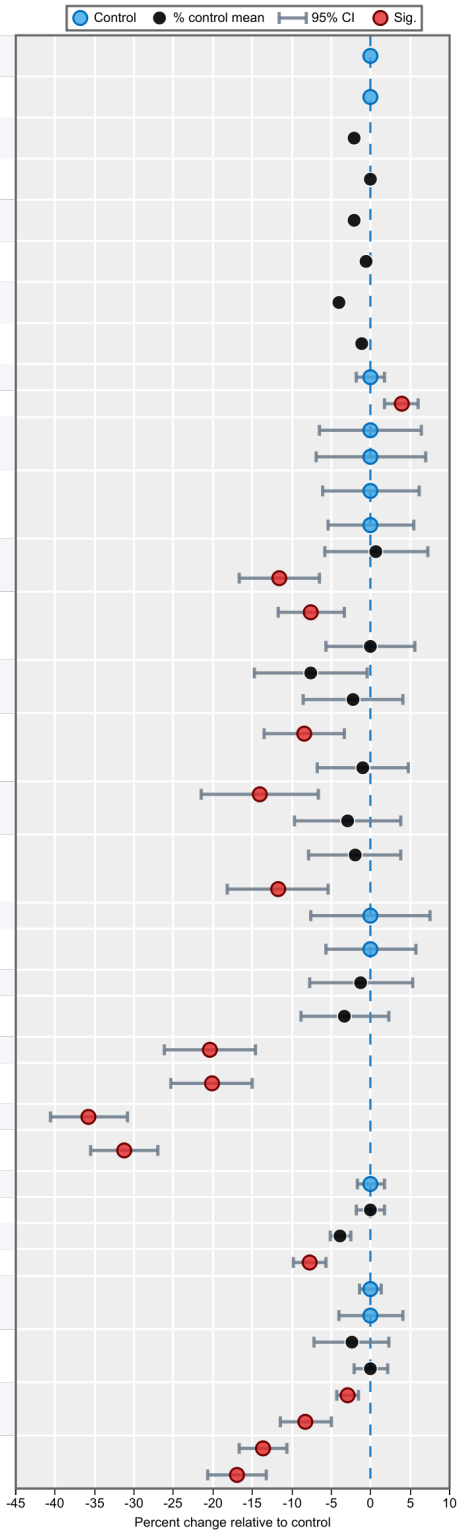


FIGURE C4-13 Data pivot of animal studies of BzBP and AGD in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-agd-dose-response/.

BzBP and Fetal Testosterone

Two studies of BzBP and effects on fetal testosterone in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. Both studies accounted for litter effects. See Figure C4-14.
- **Unexplained inconsistencies:** No downgrade. See Figure C4-15.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. Mean versus standard deviation for the studies reflects reasonable precision. See Figure C4-15.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Higher dose groups reflect a relatively large magnitude of change (about 80% in both studies). See Figure C4-15.
- **Dose-response:** Upgraded. Both studies reflect a dose response in the same dose range. See Figure C4-16.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

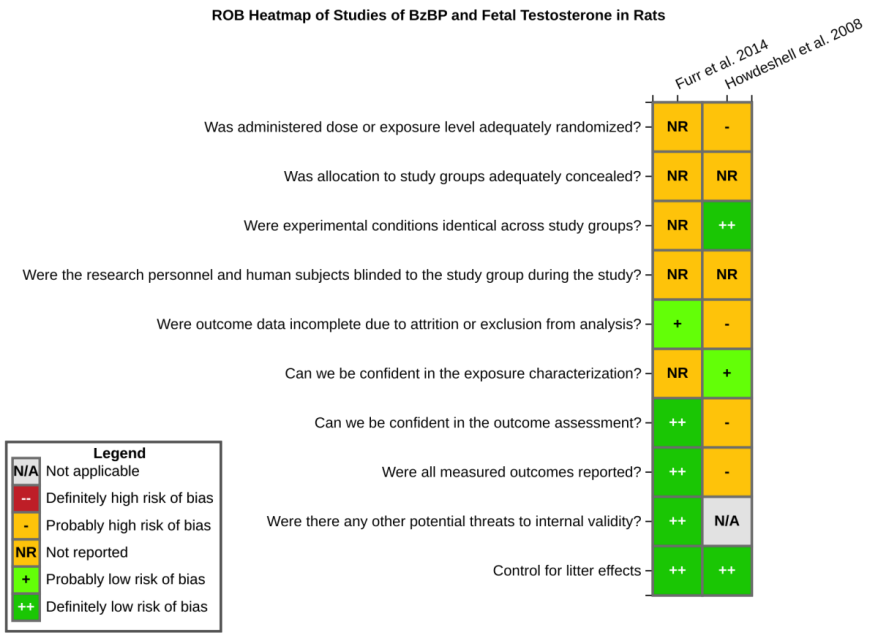


FIGURE C4-14 Risk of bias heatmap of studies of BzBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/328/.

species	dose	Endpoint	lifestage exposed	Observation time	Study
Rat	0	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	0	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation	Howdeshell et al. 2008
Rat	11	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	33	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	100	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	100	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation	Howdeshell et al. 2008
Rat	300	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	300	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation	Howdeshell et al. 2008
Rat	600	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	600	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation	Howdeshell et al. 2008
Rat	900	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation	Furr et al. 2014
Rat	900	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation	Howdeshell et al. 2008

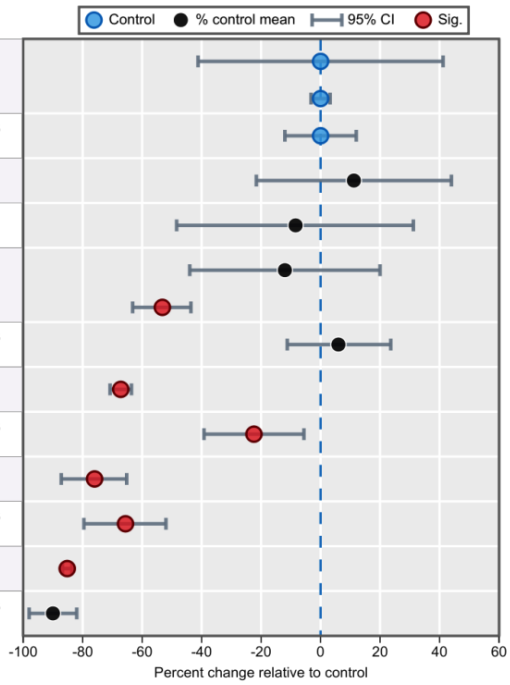


FIGURE C4-15 Data pivot of animal studies of BzBP and fetal testosterone in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-testosterone/.

species	Study	dose	Endpoint	lifestage exposed	Observation time
Rat	Furr et al. 2014	0	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	11	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	33	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	100	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	300	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	600	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Furr et al. 2014	900	Testosterone (Testes) production	GD 14 – 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	0	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation
Rat	Howdeshell et al. 2008	100	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation
Rat	Howdeshell et al. 2008	300	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation
Rat	Howdeshell et al. 2008	600	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation
Rat	Howdeshell et al. 2008	900	Testosterone (Testes) production	GD 8 - 18	GD 18 2 hr incubation

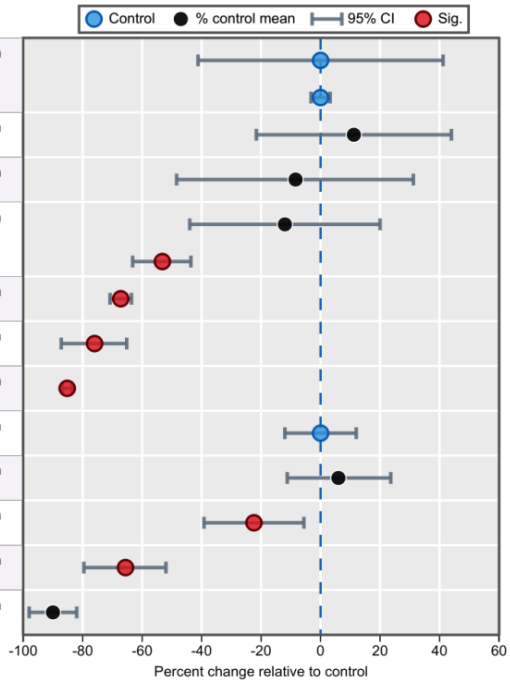


FIGURE C4-16 Data pivot of animal studies of BzBP and fetal testosterone in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-testosterone-dose-response/.

BzBP and Hypospadias

Two studies of BzBP and hypospadias in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. Both of the studies had probably high risk of bias ratings because of concerns about whether the researchers were blinded to the treatment groups and concerns about the outcome measures. One study did not control for litter effects, but it reported no hypospadias. See Figure C4-17.
- **Unexplained inconsistencies:** No downgrade. Little response seen in either study. See Figure C4-18.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. See Figure C4-18.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. Only a single hypospadias case was reported in the highest dose group in one study. See Figure C4-18.
- **Dose-response:** No upgrade. See Figure C4-18.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.
- **Rare outcome:** Because the data are limited and there were risk of bias concerns regarding the outcome measure, confidence was not upgraded for the finding of a rare effect as was done for other phthalates and hypospadias.

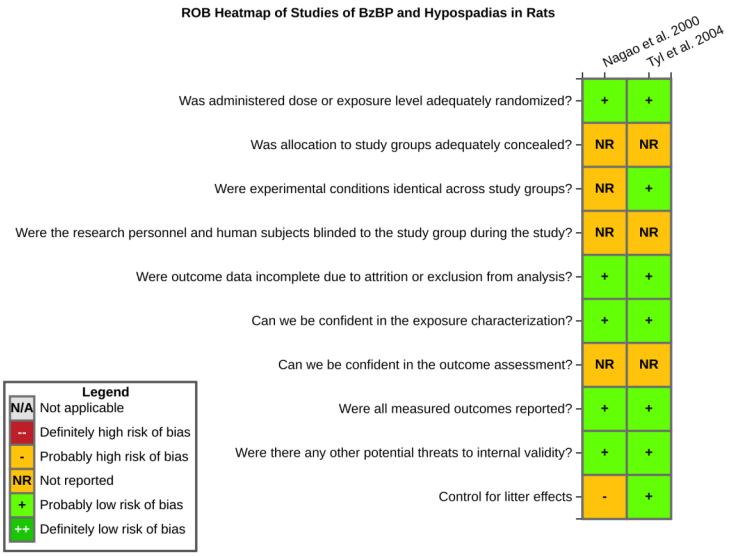


FIGURE C4-17 Risk of bias heatmap of studies of BzBP and hypospadias in rats. In HAWC: https://hawcproject.org/summary/visual/335/.

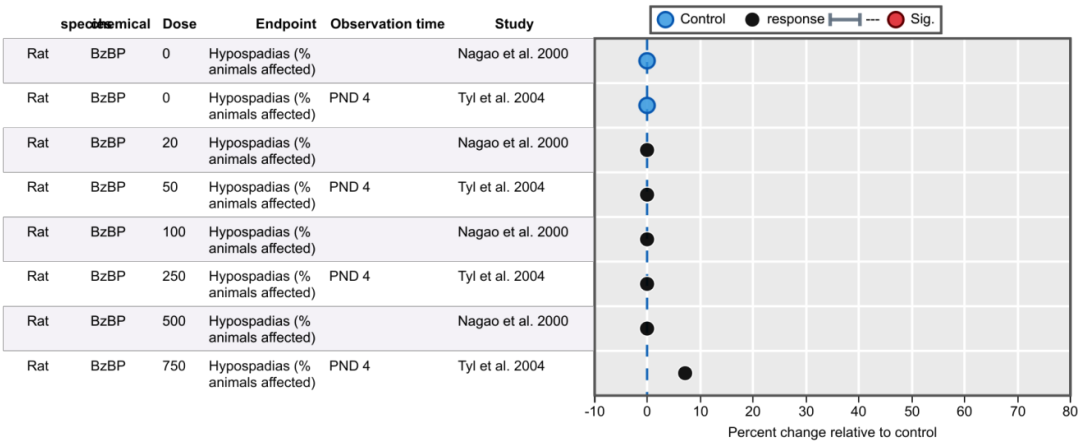


FIGURE C4-18 Data pivot of animal studies of BzBP and hypospadias (% animals affected) in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-hypospadias-animals-affected/.

The following links have additional visualizations presenting data on hypospadias in terms of the percentage of litters affected (https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-hypospadias-litters-affected/) or litter incidence (https://hawcproject.org/summary/data-pivot/assessment/351/bzbp-effect-hypospadias-litter-incidence/).

DBP and AGD

Twenty-two studies of DBP and effects on AGD in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. All studies had ratings of probably high risk of bias or definitely high risk of bias in at least one of the key issues considered, and most of the studies had multiple risk of bias issues. See Figure C4-19.
- **Unexplained inconsistencies:** No downgrade. Consistent effects observed across multiple studies. Inconsistencies could be explained by study design features. See Figure C4-20.
- **Imprecision:** No downgrade. Mean versus standard deviation for most studies reflects reasonable precision, with the exception of the study by Struve et al. (2009). See Figure C4-20.
- **Indirectness:** No downgrade

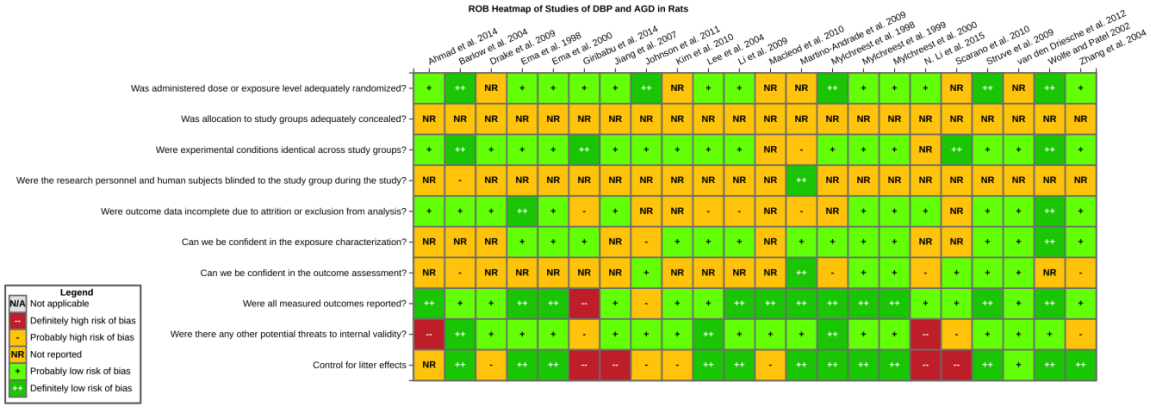


FIGURE C4-19 Risk of bias heatmap of studies of DBP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/322/.

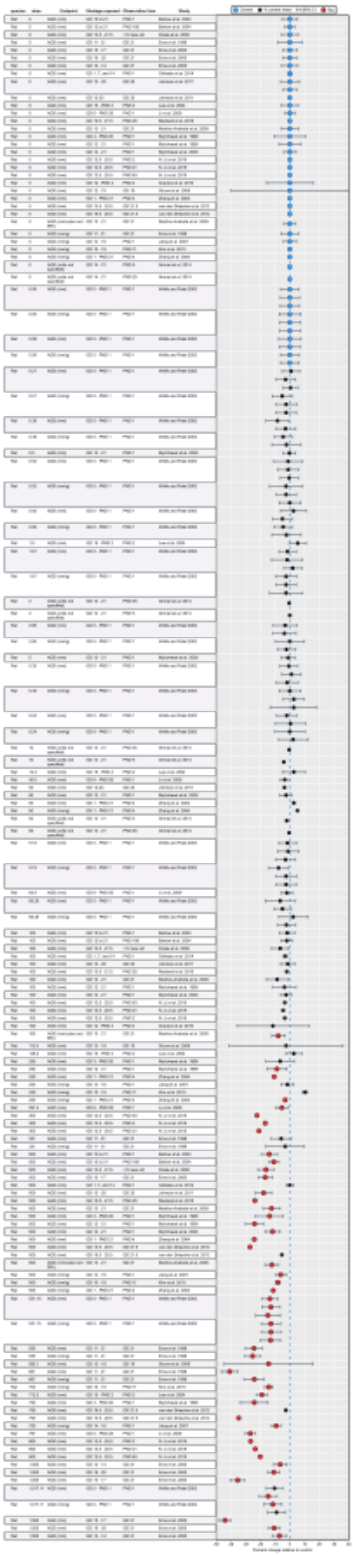


FIGURE C4-20 Data pivot of animal studies of DBP and AGD in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-agd/.

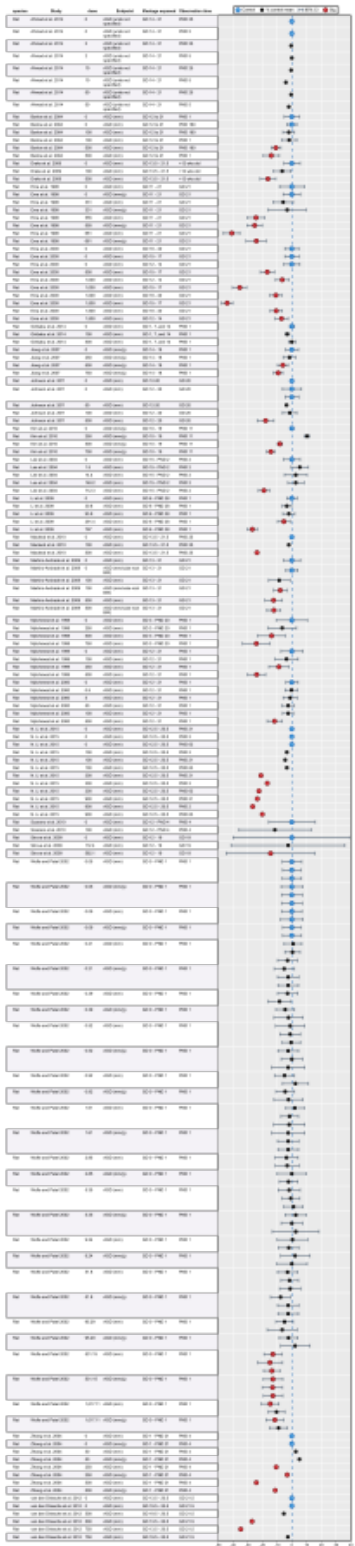


FIGURE C4-21 Data pivot of animal studies of DBP and AGD in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-agd-dose-response/.

- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. Only a few studies demonstrate a large effect (40%) even at higher doses. See Figure C4-20.
- **Dose-response:** Upgraded. Several studies reflect a dose response. See Figure C4-21.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

DBP and Fetal Testosterone

Twelve studies of DBP and effects on fetal testosterone in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure C4-22.
- **Unexplained inconsistencies:** No downgrade. Consistent effects observed across multiple studies. Inconsistencies could be explained by study design features. See Figure C4-23.
- **Indirectness:** No downgrade
- **Imprecision:** No downgrade. Mean versus standard deviation for most studies reflects reasonable precision. See Figure C4-23.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Several studies demonstrate large effects (about 80%) in high dose groups. See Figure C4-23.
- **Dose-response:** Upgraded. Several studies reflect a dose response. See Figure C4-24.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

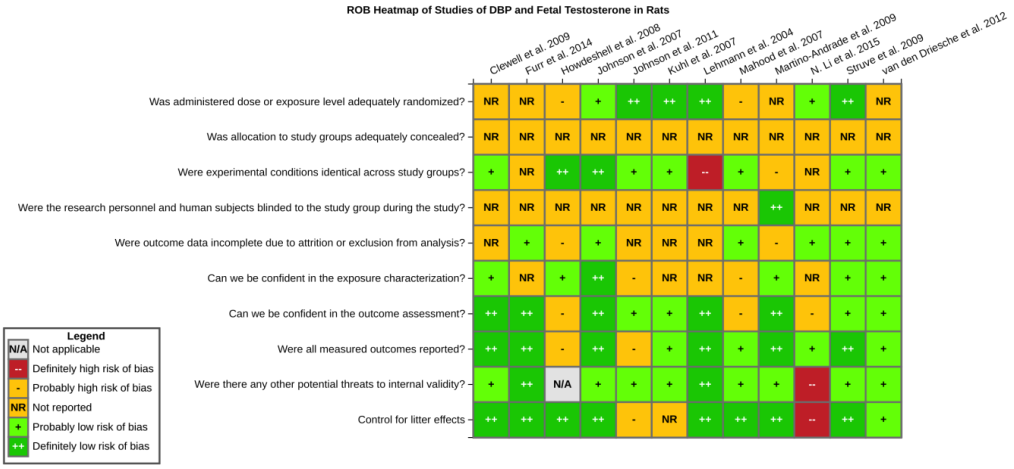


FIGURE C4-22 Risk of bias heatmap of studies of DBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/329/.

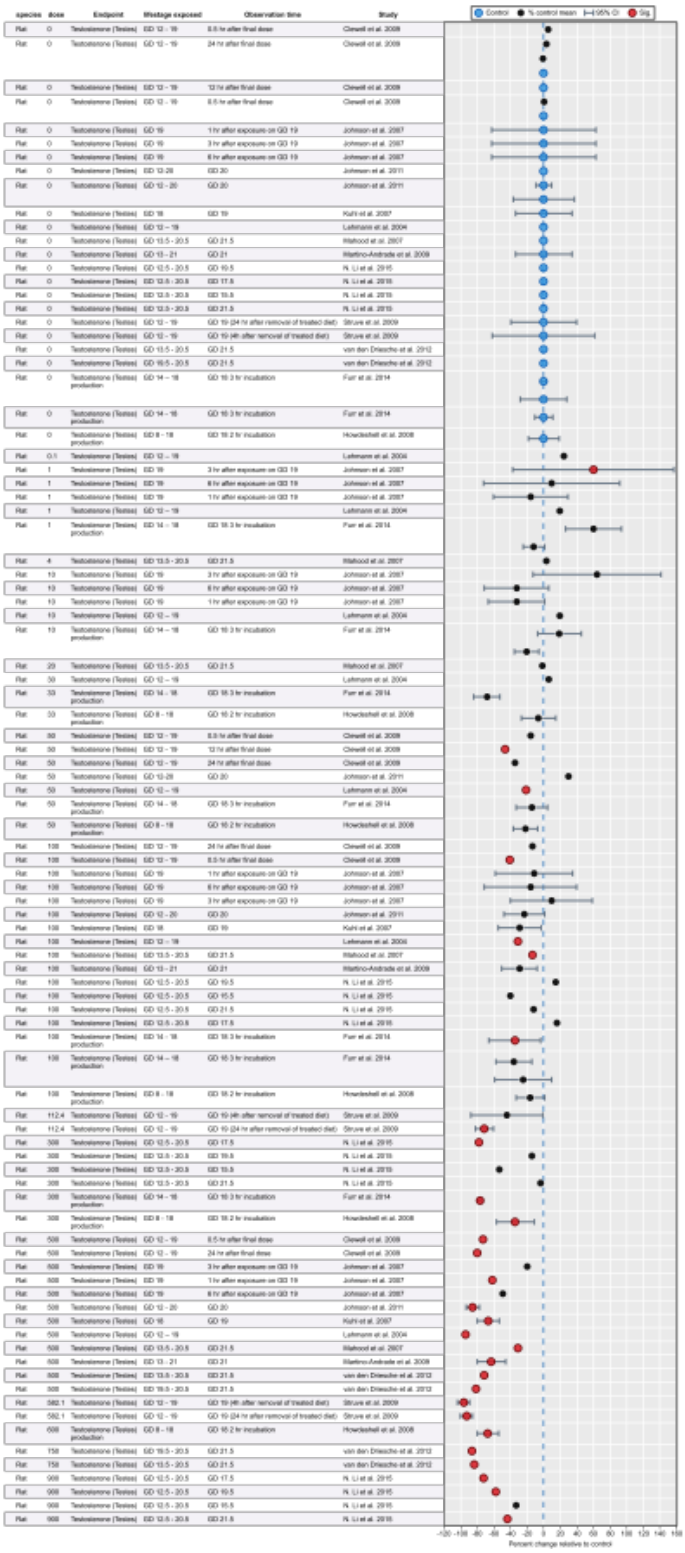


FIGURE C4-23 Data pivot of animal studies of DBP and fetal testosterone in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-testosterone/.

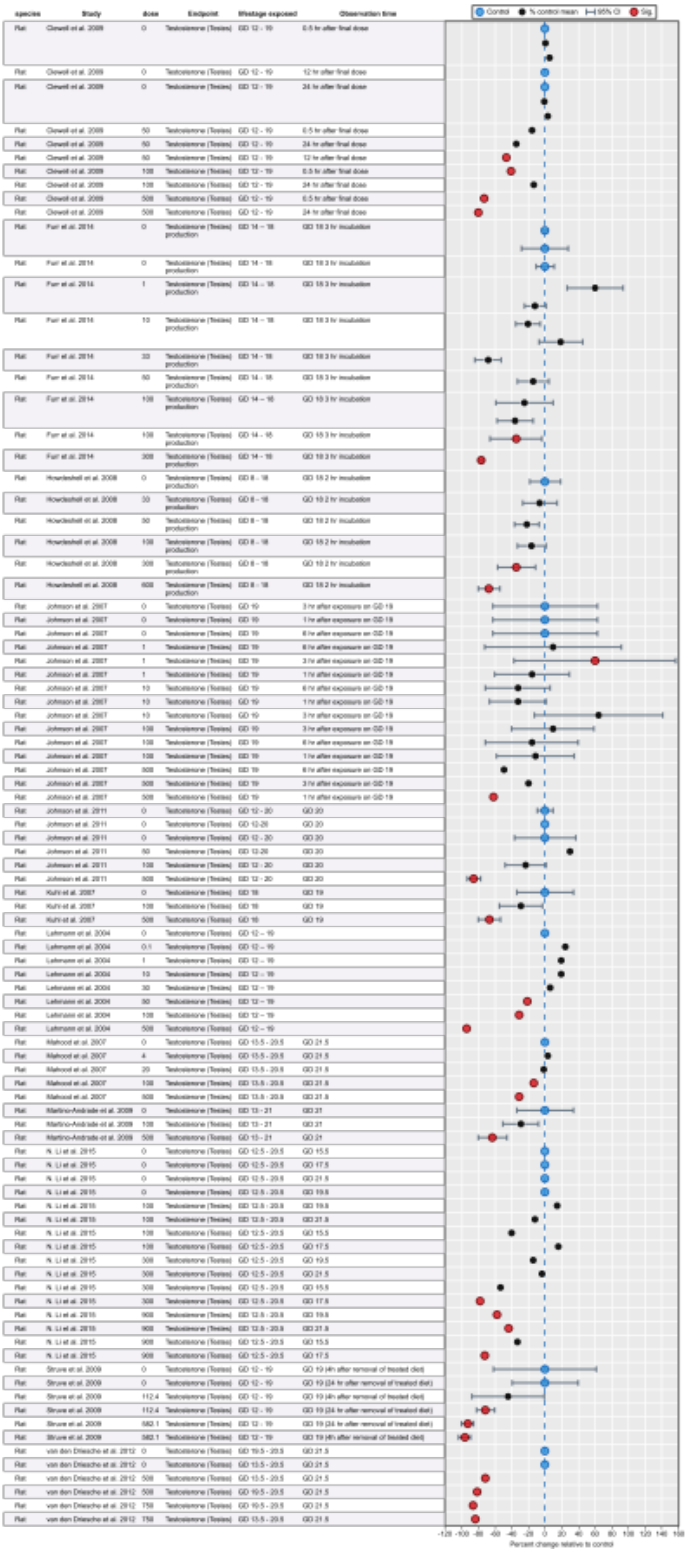


FIGURE C4-24 Data pivot of animal studies of DBP and fetal testosterone in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-testosterone-dose-response/.

DBP and Hypospadias

Eight studies of DBP and effects on hypospadias in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. Risk of bias concerns included confidence in the outcome assessment and whether the researchers were blinded to the treatment groups. See Figure C4-25.
- **Unexplained inconsistencies:** No downgrade. Incidence of hypospadias appeared to be consistent across studies within similar dose ranges. See Figure C4-26.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. No confidence intervals for incidence data, but no hypospadias in control groups. See Figure C4-26.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. High incidence (about 40%) of hypospadias was found in different studies in high dose groups. See Figure C4-26.
- **Dose-response:** Upgraded. Dose response was noted for studies and there was general agreement across studies. See Figure C4-27.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.
- **Rare outcome:** Upgraded. Background control incidence of hypospadias was reported as zero across all studies, so any positive finding was considered treatment related.

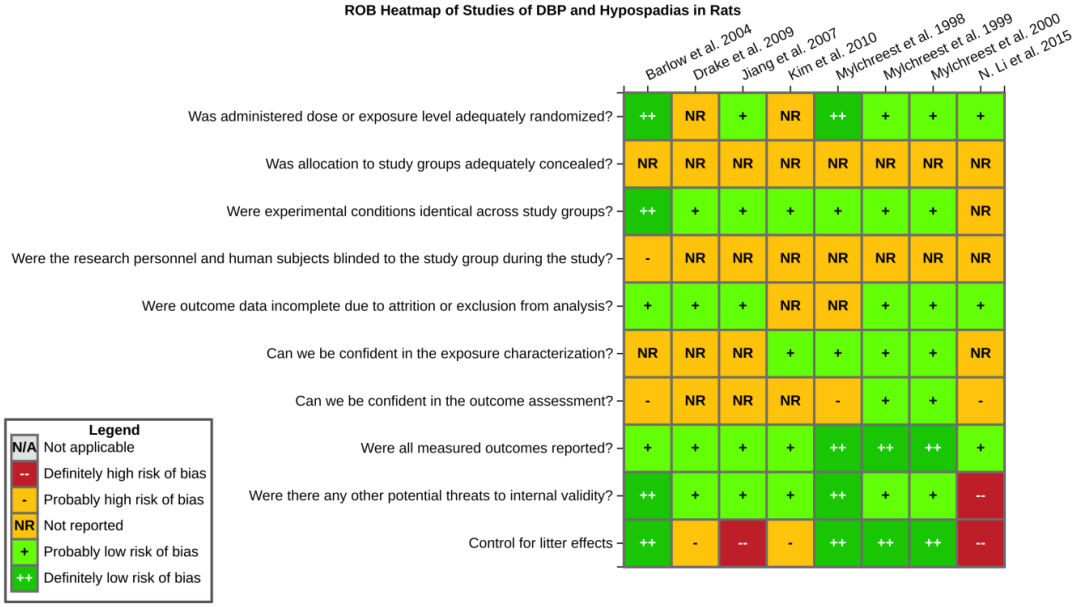


FIGURE C4-25 Risk of bias heatmap of studies of DBP and hypospadias in rats. In HAWC: https://hawcproject.org/summary/visual/338/.

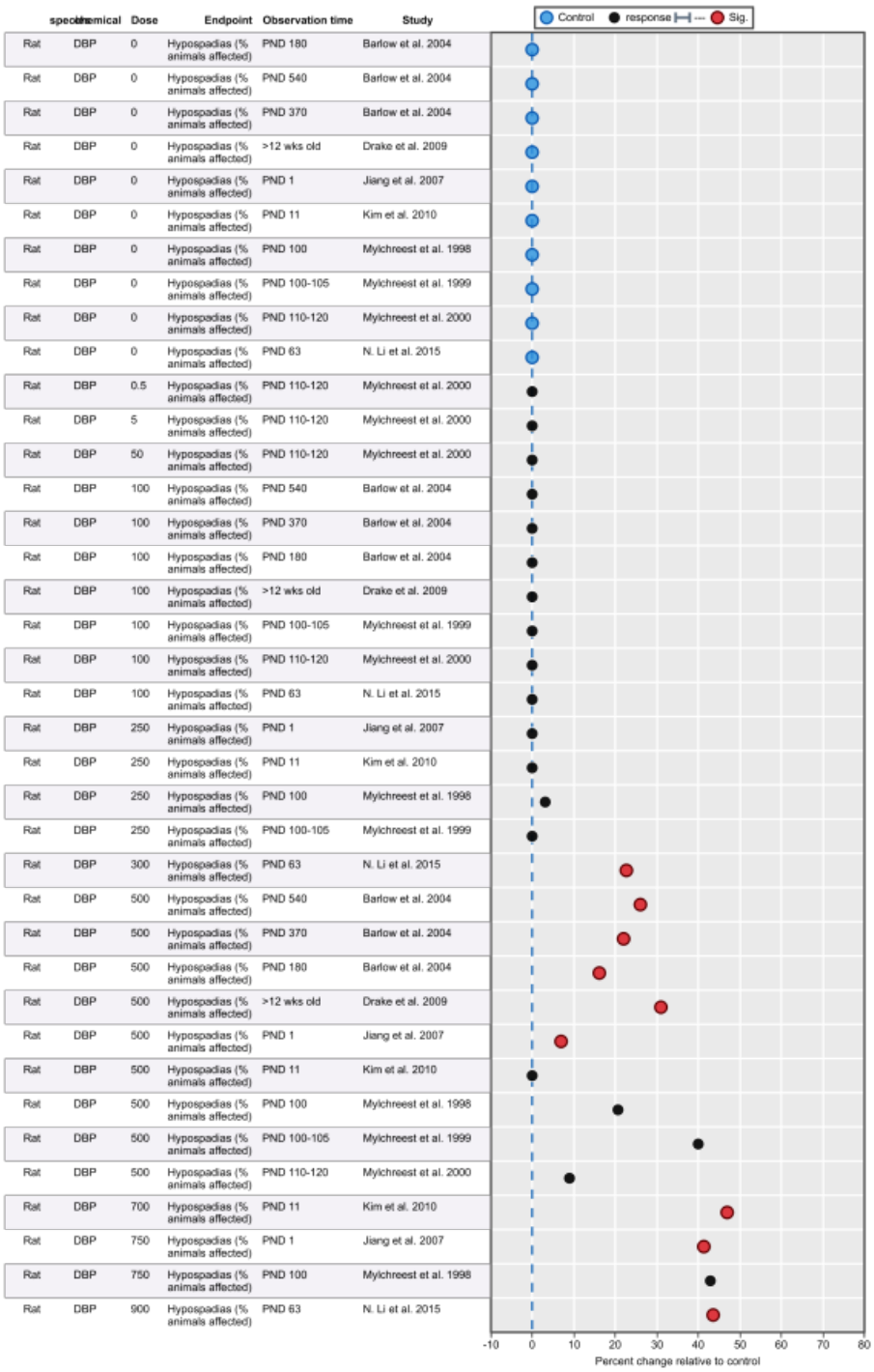
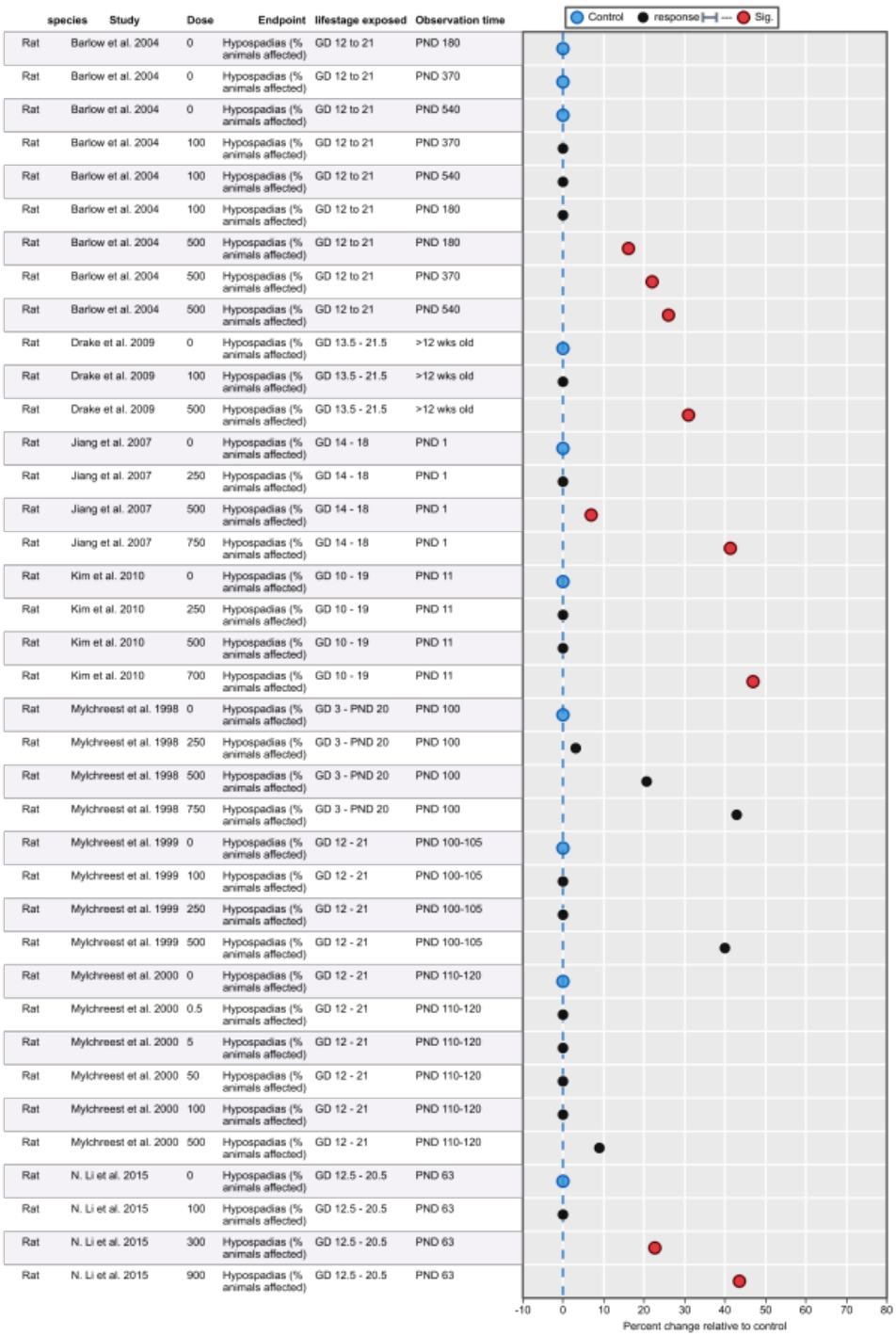


FIGURE C4-26 Data pivot of animal studies of DBP and hypospadias in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-animals-affected/.

The following links have additional visualizations presenting data on hypospadias in terms of the percentage of litters affected (https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-litters-affected/) or litter incidence (https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-litter-incidence/).



species	Study	Dose	Endpoint	lifestage exposed	Observation time
Rat	Barlow et al. 2004	0	Hypospadias (% animals affected)	GD 12 to 21	PND 180
Rat	Barlow et al. 2004	0	Hypospadias (% animals affected)	GD 12 to 21	PND 370
Rat	Barlow et al. 2004	0	Hypospadias (% animals affected)	GD 12 to 21	PND 540
Rat	Barlow et al. 2004	100	Hypospadias (% animals affected)	GD 12 to 21	PND 370
Rat	Barlow et al. 2004	100	Hypospadias (% animals affected)	GD 12 to 21	PND 540
Rat	Barlow et al. 2004	100	Hypospadias (% animals affected)	GD 12 to 21	PND 180
Rat	Barlow et al. 2004	500	Hypospadias (% animals affected)	GD 12 to 21	PND 180
Rat	Barlow et al. 2004	500	Hypospadias (% animals affected)	GD 12 to 21	PND 370
Rat	Barlow et al. 2004	500	Hypospadias (% animals affected)	GD 12 to 21	PND 540
Rat	Drake et al. 2009	0	Hypospadias (% animals affected)	GD 13.5 - 21.5	>12 wks old
Rat	Drake et al. 2009	100	Hypospadias (% animals affected)	GD 13.5 - 21.5	>12 wks old
Rat	Drake et al. 2009	500	Hypospadias (% animals affected)	GD 13.5 - 21.5	>12 wks old
Rat	Jiang et al. 2007	0	Hypospadias (% animals affected)	GD 14 - 18	PND 1
Rat	Jiang et al. 2007	250	Hypospadias (% animals affected)	GD 14 - 18	PND 1
Rat	Jiang et al. 2007	500	Hypospadias (% animals affected)	GD 14 - 18	PND 1
Rat	Jiang et al. 2007	750	Hypospadias (% animals affected)	GD 14 - 18	PND 1
Rat	Kim et al. 2010	0	Hypospadias (% animals affected)	GD 10 - 19	PND 11
Rat	Kim et al. 2010	250	Hypospadias (% animals affected)	GD 10 - 19	PND 11
Rat	Kim et al. 2010	500	Hypospadias (% animals affected)	GD 10 - 19	PND 11
Rat	Kim et al. 2010	700	Hypospadias (% animals affected)	GD 10 - 19	PND 11
Rat	Mylchreest et al. 1998	0	Hypospadias (% animals affected)	GD 3 - PND 20	PND 100
Rat	Mylchreest et al. 1998	250	Hypospadias (% animals affected)	GD 3 - PND 20	PND 100
Rat	Mylchreest et al. 1998	500	Hypospadias (% animals affected)	GD 3 - PND 20	PND 100
Rat	Mylchreest et al. 1998	750	Hypospadias (% animals affected)	GD 3 - PND 20	PND 100
Rat	Mylchreest et al. 1999	0	Hypospadias (% animals affected)	GD 12 - 21	PND 100-105
Rat	Mylchreest et al. 1999	100	Hypospadias (% animals affected)	GD 12 - 21	PND 100-105
Rat	Mylchreest et al. 1999	250	Hypospadias (% animals affected)	GD 12 - 21	PND 100-105
Rat	Mylchreest et al. 1999	500	Hypospadias (% animals affected)	GD 12 - 21	PND 100-105
Rat	Mylchreest et al. 2000	0	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	Mylchreest et al. 2000	0.5	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	Mylchreest et al. 2000	5	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	Mylchreest et al. 2000	50	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	Mylchreest et al. 2000	100	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	Mylchreest et al. 2000	500	Hypospadias (% animals affected)	GD 12 - 21	PND 110-120
Rat	N. Li et al. 2015	0	Hypospadias (% animals affected)	GD 12.5 - 20.5	PND 63
Rat	N. Li et al. 2015	100	Hypospadias (% animals affected)	GD 12.5 - 20.5	PND 63
Rat	N. Li et al. 2015	300	Hypospadias (% animals affected)	GD 12.5 - 20.5	PND 63
Rat	N. Li et al. 2015	900	Hypospadias (% animals affected)	GD 12.5 - 20.5	PND 63

FIGURE C4-27 Data pivot of animal studies of DBP and hypospadias in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-animals-affected-dose-respo/.

The following links have additional visualizations presenting data on hypospadias in terms of the percentage of litters affected (https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-litters-affected-dose-respo/) or litter incidence (https://hawcproject.org/summary/data-pivot/assessment/351/dbp-effect-hypospadias-litter-incidence-dose-respo/).

DIBP and Fetal Testosterone

Two studies of DIBP and effects on fetal testosterone in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure C4-28.
- **Unexplained inconsistencies:** No downgrade. Studies are relatively consistent. See Figure C4-29.
- **Indirectness:** No downgrade.
- **Imprecision:** Downgraded. Mean versus standard deviation reflects variable precision across studies, including overlapping error bars between control and significant treatment groups. See Figure C4-29. Meta-analysis of the data supports the downgrade (see Appendix C, Section C-6).
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

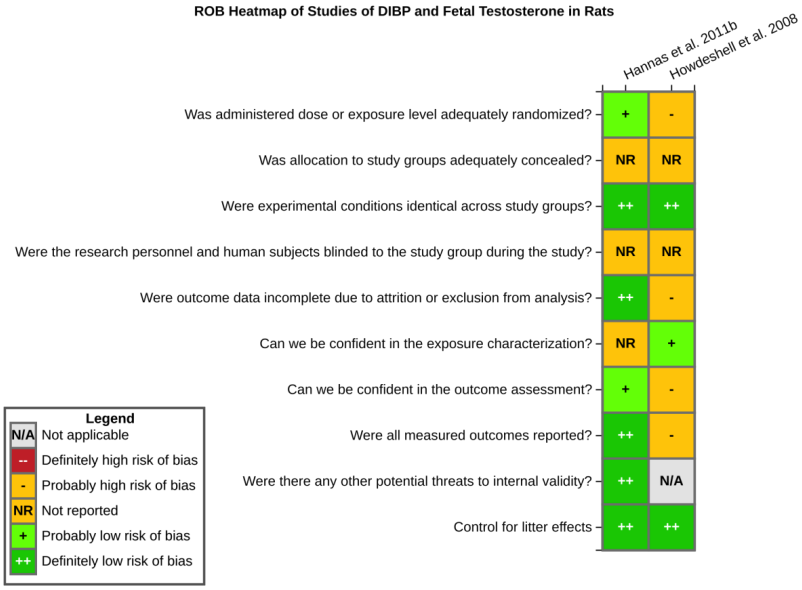


FIGURE C4-28 Risk of bias heatmap of studies of DIBP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/332/.

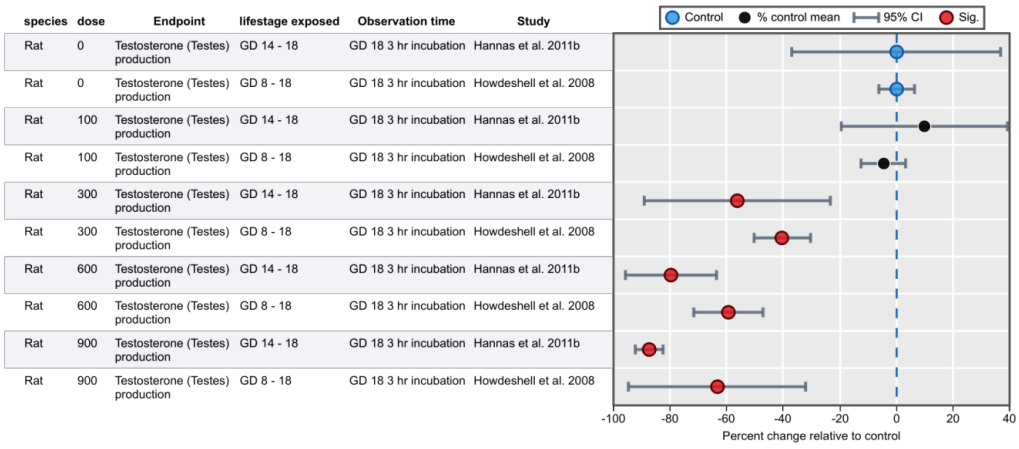
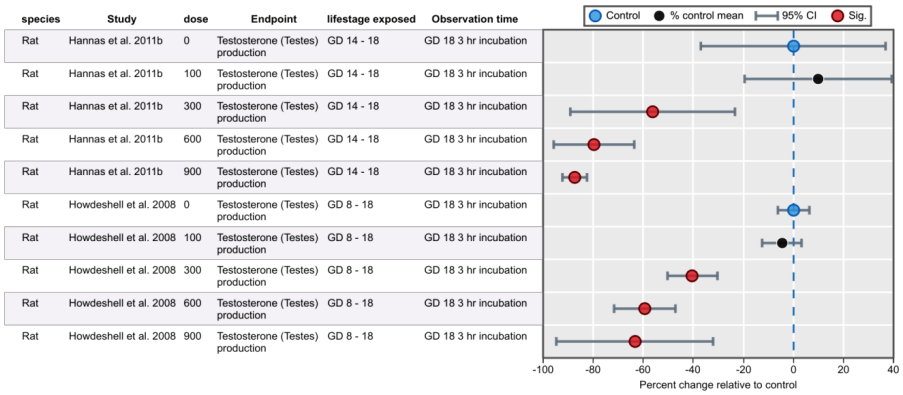


FIGURE C4-29 Data pivot of animal studies of DIBP and fetal testosterone in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dibp-effect-testosterone/.

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Consistently large effects of more than 50% are seen in both studies. See Figure C4-29.
- **Dose-response:** Upgraded. Dose response is evident in both studies. See Figure C4-30.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.



species	Study	dose	Endpoint	lifestage exposed	Observation time
Rat	Hannas et al. 2011b	0	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	100	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	300	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	600	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	900	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	0	Testosterone (Testes) production	GD 8 - 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	100	Testosterone (Testes) production	GD 8 - 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	300	Testosterone (Testes) production	GD 8 - 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	600	Testosterone (Testes) production	GD 8 - 18	GD 18 3 hr incubation
Rat	Howdeshell et al. 2008	900	Testosterone (Testes) production	GD 8 - 18	GD 18 3 hr incubation

FIGURE C4-30 Data pivot of animal studies of DIBP and fetal testosterone in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dibp-effect-testosterone-dose-response/.

DINP and AGD

Four studies of DINP and effects on AGD in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** Downgraded. Two of the studies had a probably high risk of bias rating in two key areas (whether researchers were blinded to the treatment groups or how outcomes were assessed), and one had a probably high risk of bias rating for not controlling for litter effects. See Figure C4-31.

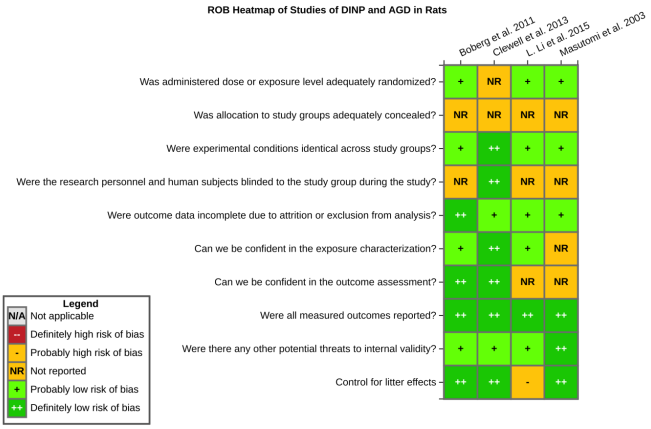


FIGURE C4-31 Risk of bias heatmap of studies of DINP and AGD in rats. In HAWC: https://hawcproject.org/summary/visual/324/.

- **Unexplained inconsistencies:** Downgraded. Responses are inconsistent across the studies and not explained by methodology or other factors. One would expect a treatment-related decrease in AGD at these dose levels given the increase in fetal testosterone; however, only one of the four studies showed a clear treatment-related decrease in AGD (Boberg et al. 2011). The study by Clewell et al. (2013) did not find decreased AGD, nor did L. Li et al. (2015), although the error was huge in that study. Masutomi et al. (2003) data appear to be equivocal; the mean is lower but the error is large. See Figure C4-32.
- **Indirectness:** No downgrade
- **Imprecision:** Downgraded. The L. Li et al. (2015) and Masutomi et al. (2003) studies had larger standard deviations than did the effect measured by Clewell et al. (2013) and Boberg et al. (2011). See Figure C4-32.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

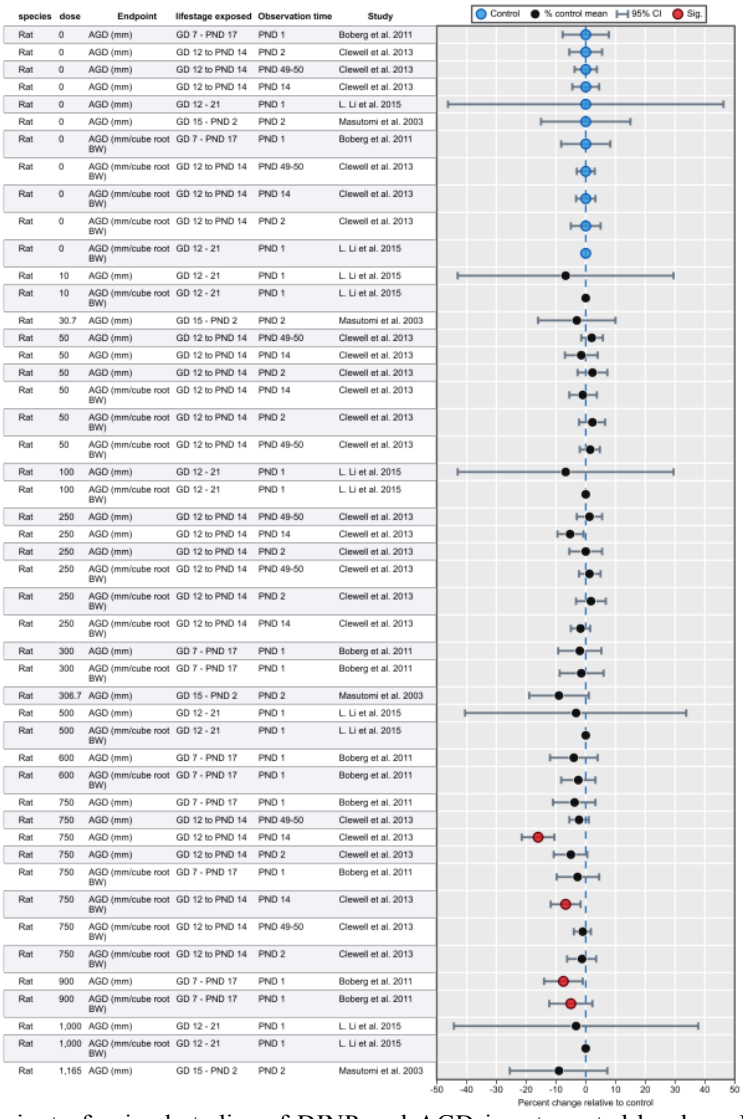


FIGURE C4-32 Data pivot of animal studies of DINP and AGD in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dinp-effect-agd/.

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. A large magnitude of effect (less than 20%) was not observed consistently across multiple studies. See Figure C4-32.
- **Dose-response:** No upgrade. Dose response is not consistent across studies. Only two of four studies show a dose response. The data from Clewell et al. (2013) are internally inconsistent; no effect at PND 2 or PND 49-50, but a statistically-identified decrease in AGD was found at PND 14. See Figure C4-33.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

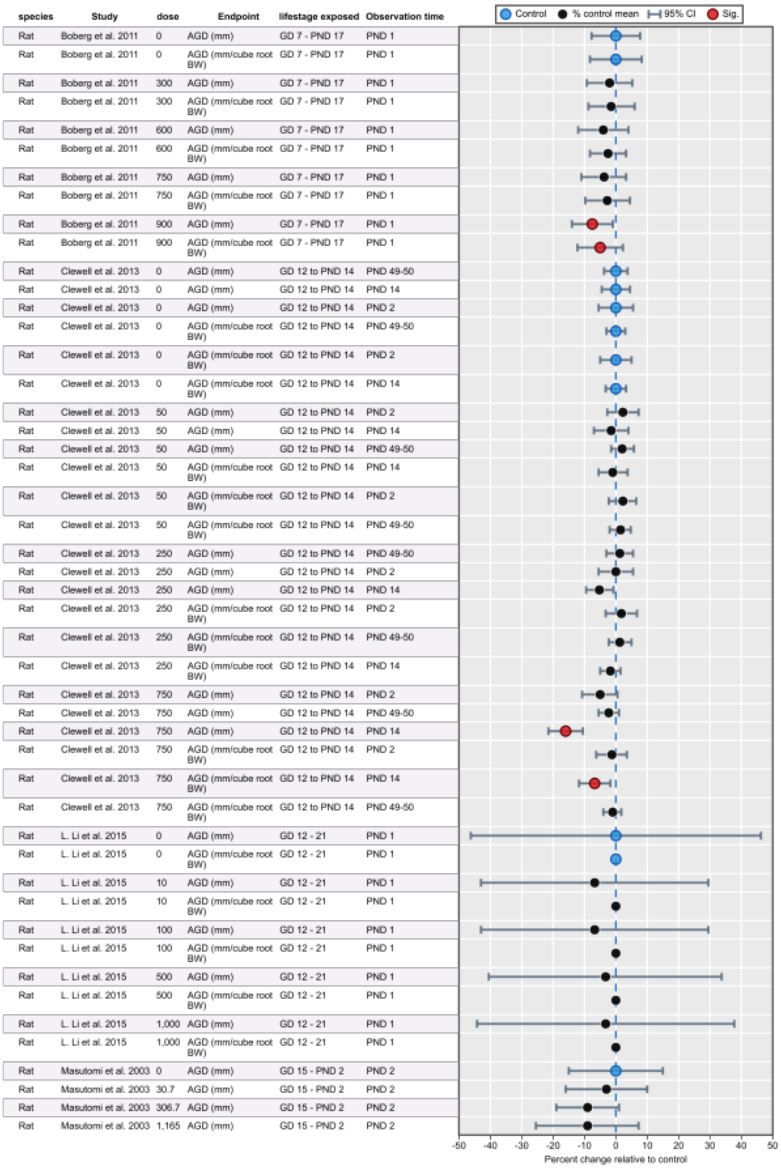


FIGURE C4-33 Data pivot of animal studies of DINP and AGD in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dinp-effect-agd-dose-response/.

DINP and Fetal Testosterone

Four studies of DINP and effects on fetal testosterone in rats were available.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure C4-34.
- **Unexplained inconsistencies:** No downgrade. Inconsistencies can be explained by study design features (exposure window) and differences in measurements (testosterone in plasma is different from testosterone in the testes). See Figure C4-35.
- **Indirectness:** No downgrade.
- **Imprecision:** Downgraded. Mean versus standard deviation reflects variable precision across studies, particularly testosterone production and testosterone measurements in plasma. See Figure C4-35. Meta-analysis of the data also supported a downgrade (see Appendix C, Section C-6).
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgraded. Studies show large effects of more than 50% (see Figure C4-35), and meta-analysis of the data found an overall effect that was large in magnitude (see Appendix C, Section C-6).
- **Dose-response:** Upgrade. Dose response is evident in most studies, although not statistically significant in most cases. See Figure C4-36.
- **Residual confounding:** Not applicable.
- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.

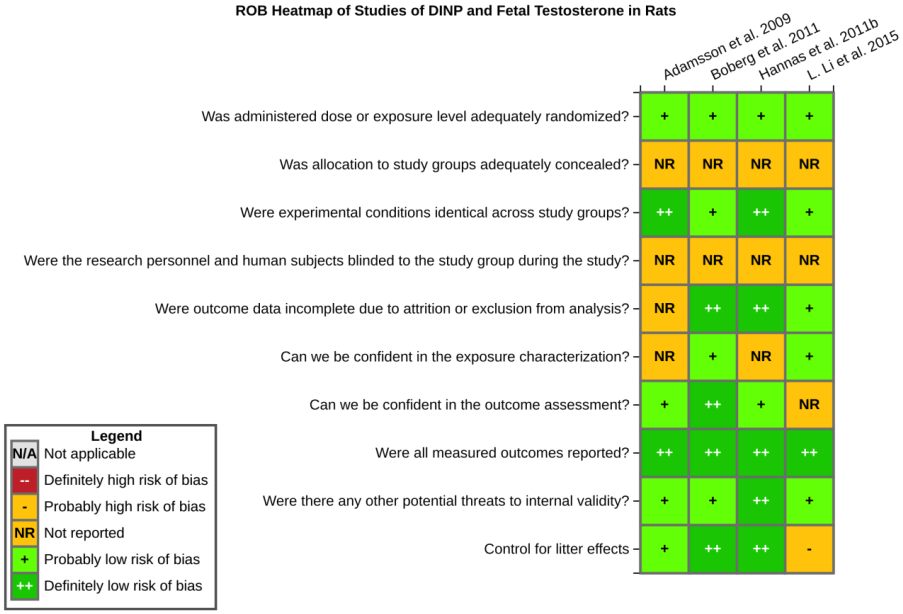


FIGURE C4-34 Risk of bias heatmap of studies of DINP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/333/.

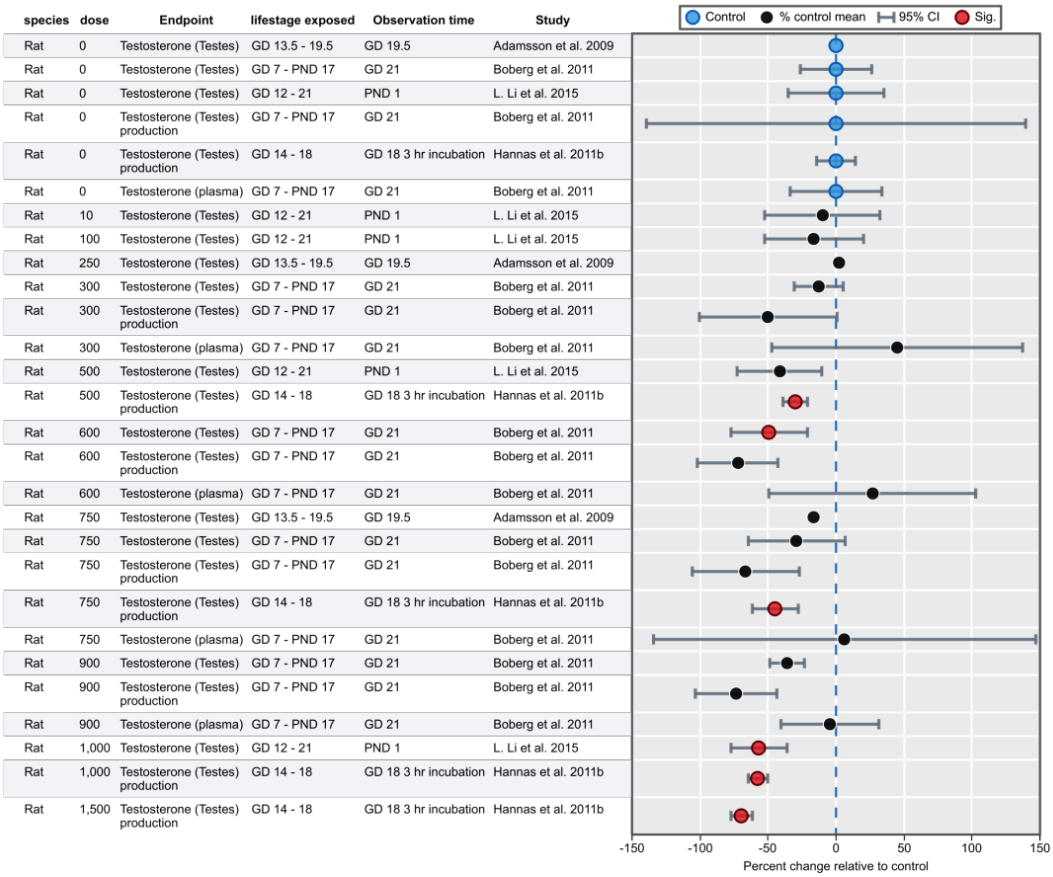


FIGURE C4-35 Data pivot of animal studies of DINP and fetal testosterone in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dinp-effect-testosterone/.

DPP and Fetal Testosterone

Four studies of DPP and effects on fetal testosterone in rats were available.

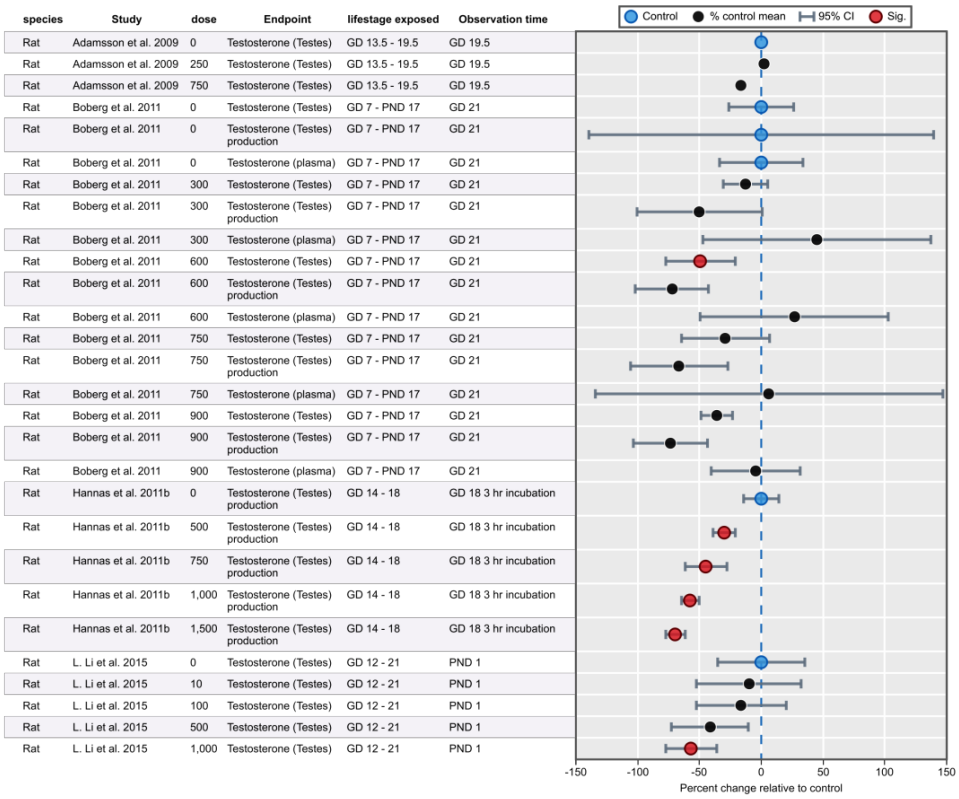
Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure C4-37.
- **Unexplained inconsistencies:** No downgrade. Data are relatively consistent across studies, and inconsistencies can be explained by study design or measurement features (incubation time). See Figure C4-38.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade. Mean versus standard deviation reflects reasonable precision across studies. See Figure C4-38.
- **Publication bias:** No downgrade (see Appendix C, Section C-3).

Factors Considered for Upgrading Confidence

- **Large magnitude:** Upgrade. Consistently large effects of more than 60% are seen in several studies within the same dose ranges. See Figure C4-38.
- **Dose-response:** Upgrade. Dose response is evident in most studies. See Figure C4-39.
- **Residual confounding:** Not applicable.

- **Cross-species consistency:** No upgrade. Only studies in rats were available so cross-species consistency could not be evaluated.



species	Study	dose	Endpoint	lifestage exposed	Observation time
Rat	Adamsson et al. 2009	0	Testosterone (Testes)	GD 13.5 - 19.5	GD 19.5
Rat	Adamsson et al. 2009	250	Testosterone (Testes)	GD 13.5 - 19.5	GD 19.5
Rat	Adamsson et al. 2009	750	Testosterone (Testes)	GD 13.5 - 19.5	GD 19.5
Rat	Boberg et al. 2011	0	Testosterone (Testes)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	0	Testosterone (Testes) production	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	0	Testosterone (plasma)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	300	Testosterone (Testes)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	300	Testosterone (Testes) production	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	300	Testosterone (plasma)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	600	Testosterone (Testes)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	600	Testosterone (Testes) production	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	600	Testosterone (plasma)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	750	Testosterone (Testes)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	750	Testosterone (Testes) production	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	750	Testosterone (plasma)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	900	Testosterone (Testes)	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	900	Testosterone (Testes) production	GD 7 - PND 17	GD 21
Rat	Boberg et al. 2011	900	Testosterone (plasma)	GD 7 - PND 17	GD 21
Rat	Hannas et al. 2011b	0	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	500	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	750	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	1,000	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	Hannas et al. 2011b	1,500	Testosterone (Testes) production	GD 14 - 18	GD 18 3 hr incubation
Rat	L. Li et al. 2015	0	Testosterone (Testes)	GD 12 - 21	PND 1
Rat	L. Li et al. 2015	10	Testosterone (Testes)	GD 12 - 21	PND 1
Rat	L. Li et al. 2015	100	Testosterone (Testes)	GD 12 - 21	PND 1
Rat	L. Li et al. 2015	500	Testosterone (Testes)	GD 12 - 21	PND 1
Rat	L. Li et al. 2015	1,000	Testosterone (Testes)	GD 12 - 21	PND 1

FIGURE C4-36 Data pivot of animal studies of DINP and fetal testosterone in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dinp-effect-testosterone-dose-response/.

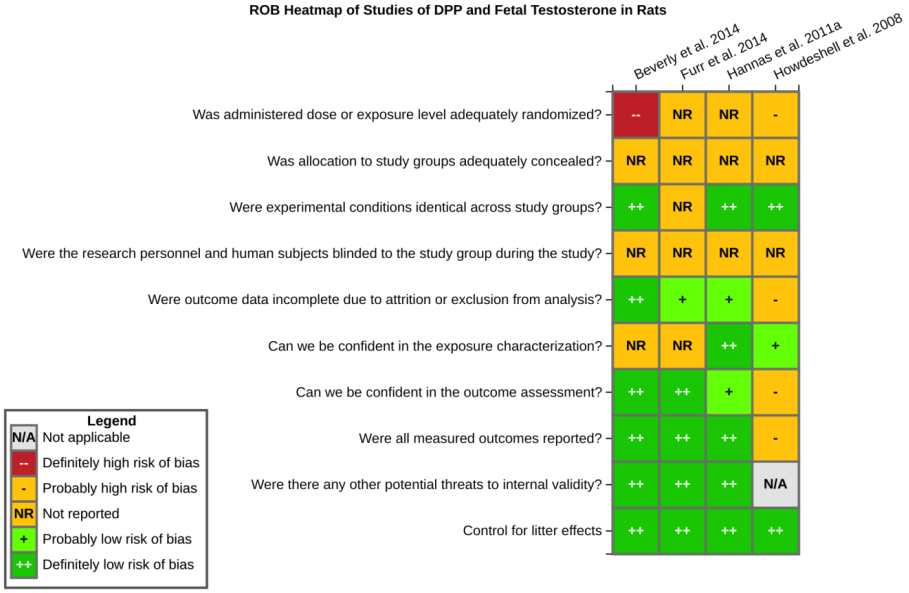


FIGURE C4-37 Risk of bias heatmap of studies of DPP and fetal testosterone in rats. In HAWC: https://hawcproject.org/summary/visual/334/.

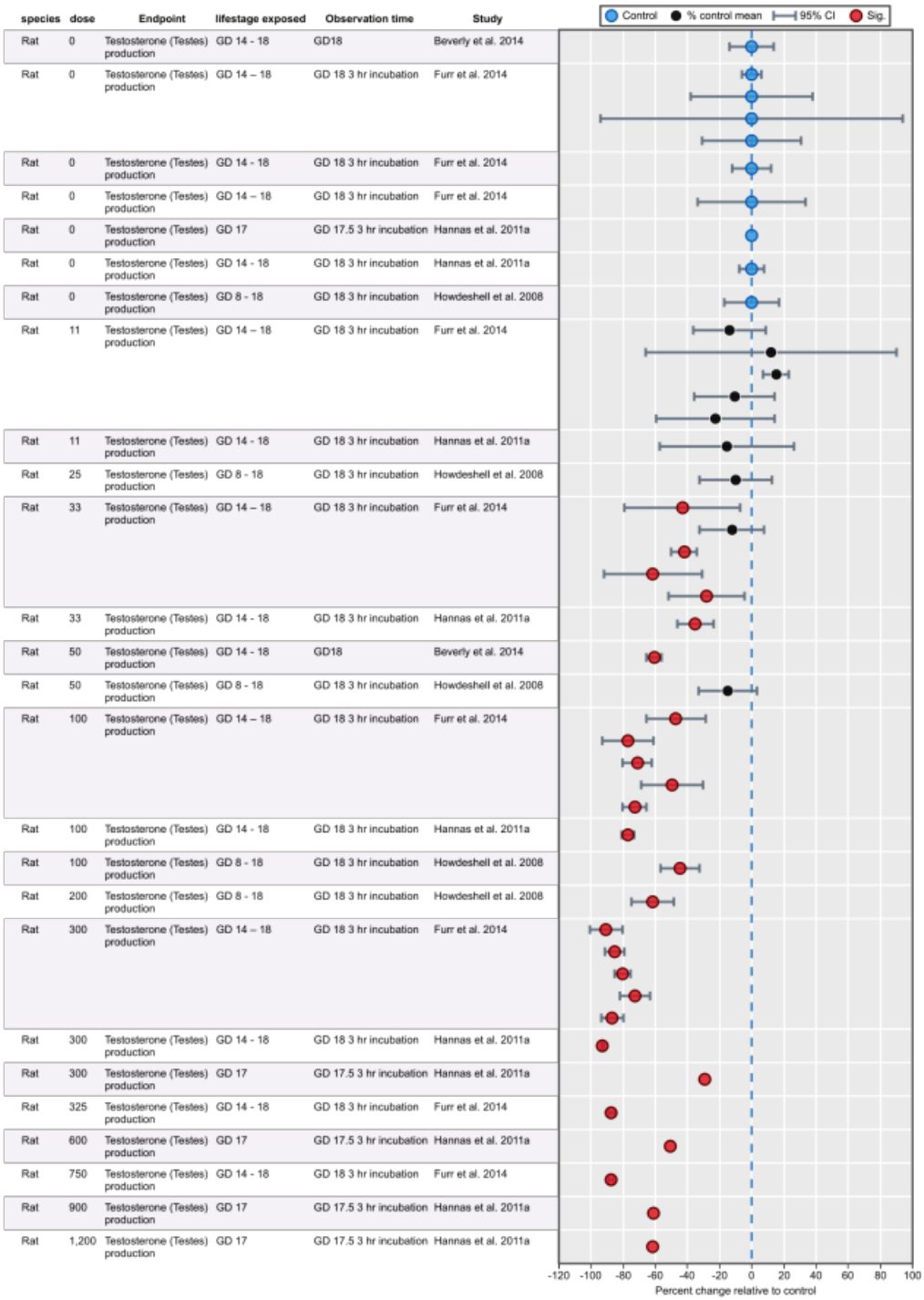


FIGURE C4-38 Data pivot of animal studies of DPP and fetal testosterone in rats sorted by dose. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dpp-effect-testosterone/.

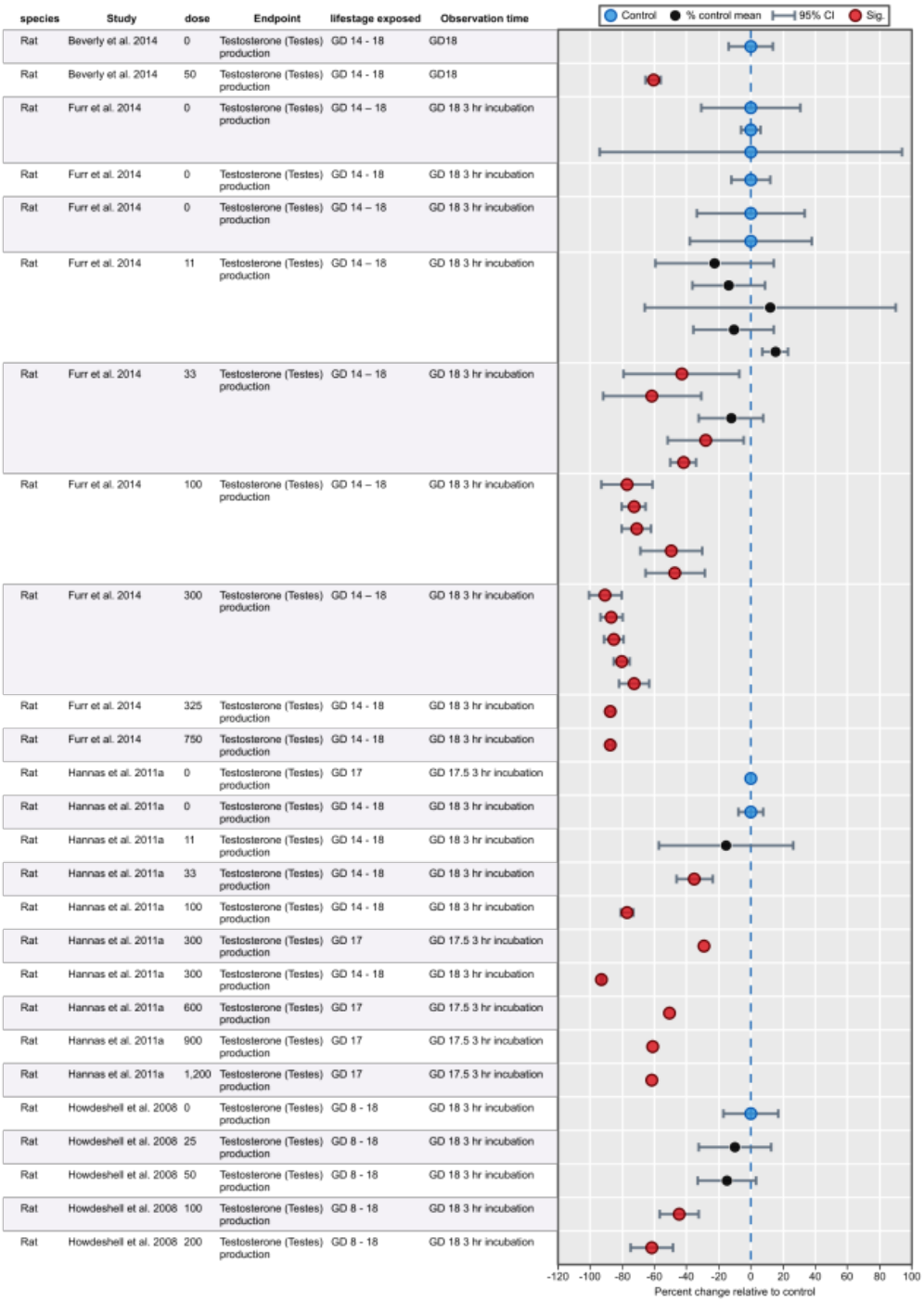


FIGURE C4-39 Data pivot of animal studies of DPP and fetal testosterone in rats sorted by study. In HAWC: https://hawcproject.org/summary/data-pivot/assessment/351/dpp-effect-testosterone-dose-response/.

SECTION C-5

Supporting Information for the Meta-analyses of Studies of DEHP

META-ANALYSES OF RAT STUDIES ON DEHP AND AGD

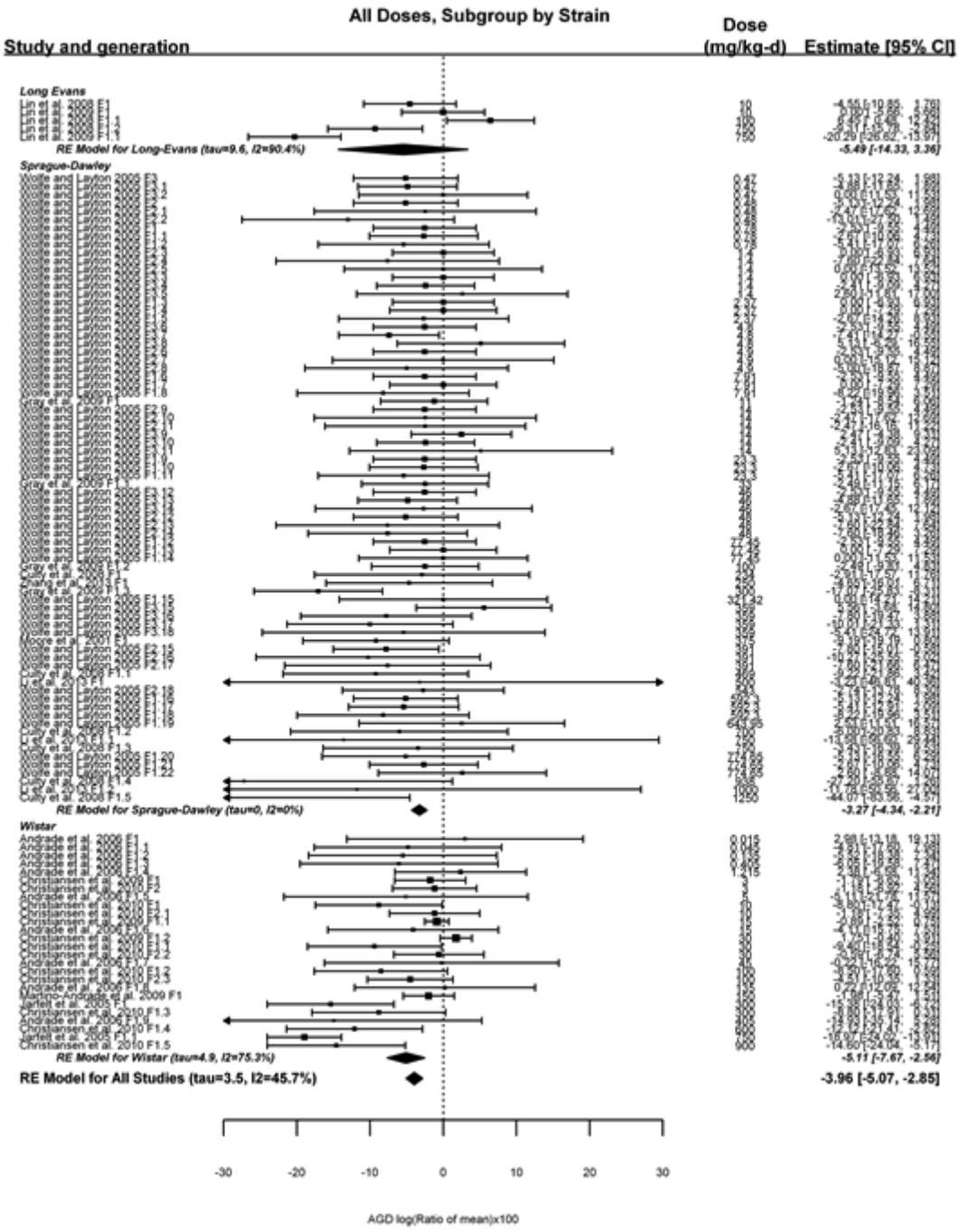


FIGURE C5-1 Results of meta-analyses of studies on DEHP and AGD in different strains of rat using the random effects model.

TABLE C5-1 Subgrouping Analyses of Rat Studies on DEHP and AGD

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Long Evans									
Rat DEHP LE Overall	Intrcpt	-5.49	-14.33	3.36	0.224	9.59	90.38	0.000	45.84
Rat DEHP LE Trend in log10 dose	log10(dose)	-6.46	-16.21	3.29	0.194	8.78	88.76	0.000	51.91
Rat DEHP LE Linear in dose100	dose100	-1.90	-3.04	-0.77	0.001	5.21	73.98	0.004	41.81*
Sprague-Dawley									
Rat DEHP SD Overall	Intrcpt	-3.27	-4.34	-2.21	0.000	0.00	0.00	0.988	448.47
Rat DEHP SD Trend in log10 dose	log10(dose)	-0.92	-1.96	0.12	0.083	0.00	0.00	0.993	442.56*
Rat DEHP SD Linear-Quadratic in dose100	dose100	-2.40	-3.78	-1.01	0.001	0.00	0.00	0.899	452.31
	I(dose100^2)	0.22	0.01	0.43	0.036				
Wistar									
Rat DEHP W Overall	Intrcpt	-5.11	-7.67	-2.56	0.000	4.94	75.25	0.000	168.41
Rat DEHP W Trend in log10 dose	log10(dose)	-3.14	-5.21	-1.06	0.003	3.94	65.66	0.000	157.70
Rat DEHP W Linear-Quadratic in dose100	dose100	-3.58	-5.57	-1.59	0.000	1.38	22.04	0.386	143.73*
	I(dose100^2)	0.18	-0.09	0.45	0.201				

*Indicates the lowest AICc for each strain.

TABLE C5-2 Overall Analyses and Sensitivity Analyses of Rat Studies of DEHP and AGD Without Strain Subgrouping

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-3.96	-5.07	-2.85	0.000	3.48	45.74	0.000	680.05
Trend in log10(dose)	log10(dose)	-1.97	-2.98	-0.96	0.000	3.00	38.46	0.000	662.96
Linear in dose100	dose100	-1.55	-1.86	-1.24	0.000	1.97	22.05	0.124	659.46
Linear-Quadratic in dose100	dose100	-2.11	-3.30	-0.91	0.001	2.03	22.94	0.117	654.59*
	I(dose100^2)	0.08	-0.09	0.25	0.337				
Sensitivity Analyses									
Overall minus Christiansen et al. 2009	intrcpt	-4.23	-5.37	-3.09	0.000	3.43	37.58	0.001	662.79
Overall minus Christiansen et al. 2010	intrcpt	-3.69	-4.88	-2.50	0.000	3.53	46.25	0.000	618.76
Overall minus Culty et al. 2008	intrcpt	-3.86	-4.99	-2.73	0.000	3.51	47.34	0.000	634.25
Overall minus Lin et al. 2008	intrcpt	-4.02	-5.13	-2.91	0.000	3.33	43.19	0.000	658.78
Overall minus Gray et al. 2009	intrcpt	-3.89	-5.02	-2.76	0.000	3.43	45.13	0.000	653.44
Overall minus Lin et al. 2009	intrcpt	-3.72	-4.78	-2.67	0.000	2.97	37.79	0.001	656.63
Overall minus Li et al. 2013	intrcpt	-3.95	-5.06	-2.84	0.000	3.49	46.49	0.000	655.79

Overall minus Jarfelt et al. 2005	intrcpt	-3.44	-4.44	-2.43	0.000	2.59	31.65	0.012	650.98
Overall minus Moore et al. 2001	intrcpt	-3.91	-5.03	-2.80	0.000	3.48	45.89	0.000	673.84
Overall minus Zhang et al. 2013	intrcpt	-3.96	-5.07	-2.84	0.000	3.50	46.16	0.000	674.38
Overall minus Andrade et al. 2006	intrcpt	-4.04	-5.20	-2.89	0.000	3.59	48.97	0.000	616.81
Overall minus Martino-Andrade et al. 2009	intrcpt	-4.01	-5.13	-2.88	0.000	3.53	45.56	0.000	675.20
Overall minus Wolfe and Layton 2005	intrcpt	-5.59	-7.72	-3.45	0.000	5.37	72.75	0.000	314.53
Highest Doses-Overall	intrcpt	-8.08	-12.31	-3.86	0.000	7.26	81.76	0.000	129.90
Highest Doses-Linear in dose100	dose100	-1.87	-2.45	-1.30	0.000	4.22	65.26	0.004	120.97
Highest Doses-Trend in log10(dose)	log10(dose)	-11.44	-19.03	-3.86	0.003	5.21	63.51	0.001	120.83
Highest Doses-Linear-Quadratic in dose100	dose100	-1.34	-3.53	0.85	0.232	4.46	67.35	0.004	117.23
	I(dose100^2)	-0.07	-0.37	0.22	0.623				

*Indicates the lowest AICc.

TABLE C5-3 Benchmark Dose Estimates for DEHP and AGD in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
All strains Linear in dose100	-5.1	332	276	415
All strains Linear-Quadratic in dose100	-5.1	271	178	418
LE Linear in dose100	-5.1	268	168	659
SD Linear-Quadratic in dose100	-5.1	294	166	NA
W Linear-Quadratic in dose100	-5.1	154	99	282

Benchmark dose estimates were calculated for an effect size of 5%. The benchmark dose was calculated using the linear or linear-quadratic model, with the model selection based on the lowest AIC (including correction for small sample size). The benchmark dose was only calculated for the "fixed effect"—the estimated mean response across studies.

TABLE C5-4 Overall Analyses and Sensitivity Analyses of Mouse Studies of DEHP and AGD

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-1.57	-4.61	1.47	0.310	3.94	82.39	0.000	68.84
Trend in log10(dose)	log10(dose)	-1.77	-2.71	-0.83	0.000	1.60	40.12	0.095	60.64
Linear in dose100	dose100	-2.03	-3.51	-0.55	0.007	2.81	69.81	0.005	64.57
Linear-Quadratic in dose100	dose100	-5.71	-7.15	-4.27	0.000	0.00	0.00	0.185	59.68*
	I(dose100^2)	0.96	0.42	1.49	0.000				
Sensitivity Analyses									
Overall minus Do et al. 2012	intrcpt	-4.48	-7.12	-1.85	0.001	2.48	80.24	0.000	37.71
Overall minus Pocar et al. 2012	intrcpt	-1.08	-5.32	3.16	0.617	5.10	85.14	0.000	59.53
Overall minus Liu et al. 2008	intrcpt	0.31	-2.38	3.00	0.821	2.19	38.54	0.186	47.17
Highest Doses-Overall	intrcpt	-2.27	-5.10	0.55	0.115	0.00	0.00	0.319	28.25
Highest Doses-Linear in dose100	dose100	-1.01	-2.65	0.63	0.228	1.14	22.50	0.195	27.46

*Indicates the lowest AICc.

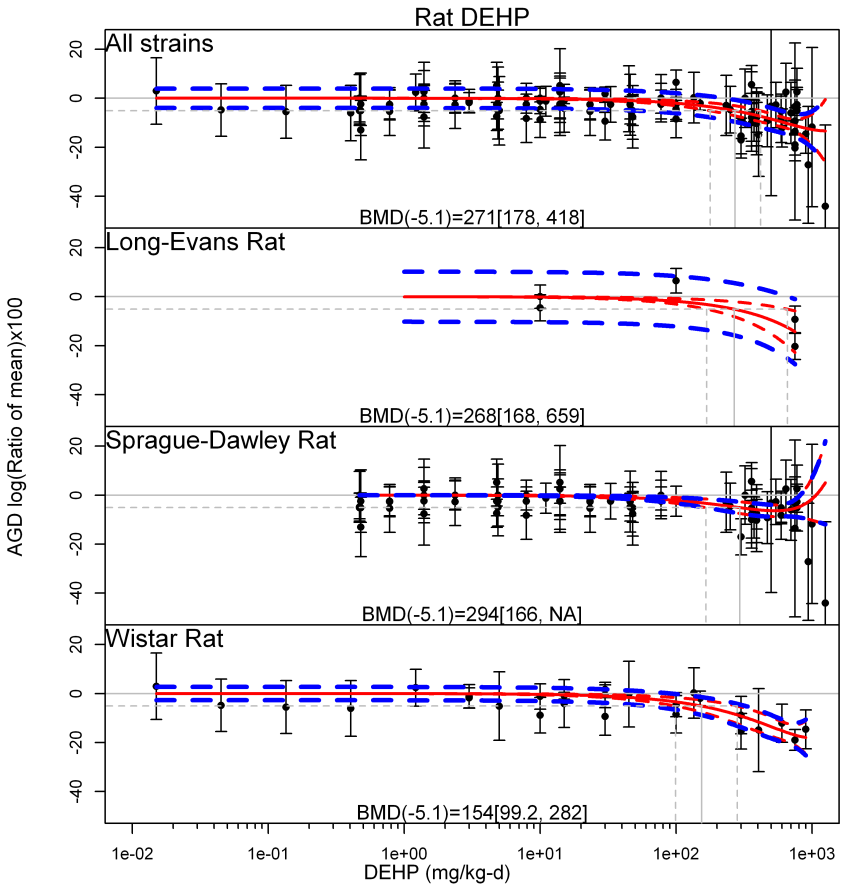


FIGURE C5-2 Benchmark dose estimates from rat studies of DEHP and AGD.

META-ANALYSES OF MOUSE STUDIES ON DEHP AND AGD

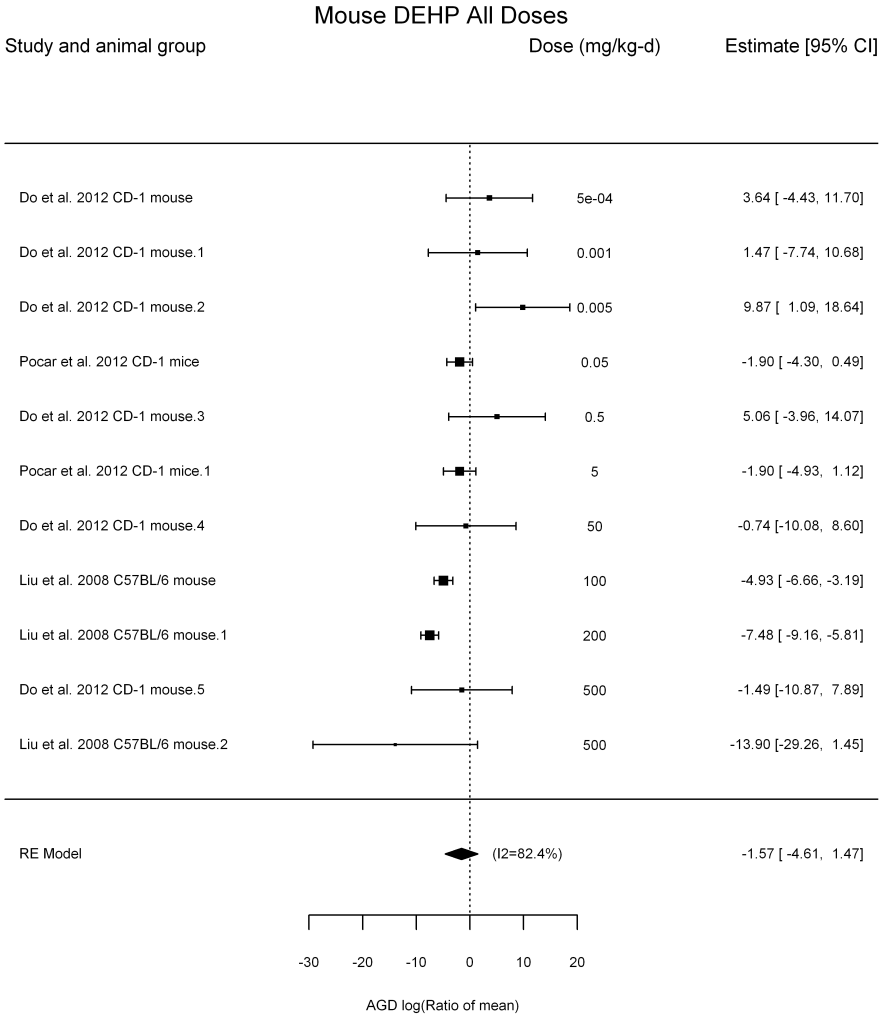


FIGURE C5-3 Results of meta-analyses of studies on DEHP and AGD in mice using the random effects model.

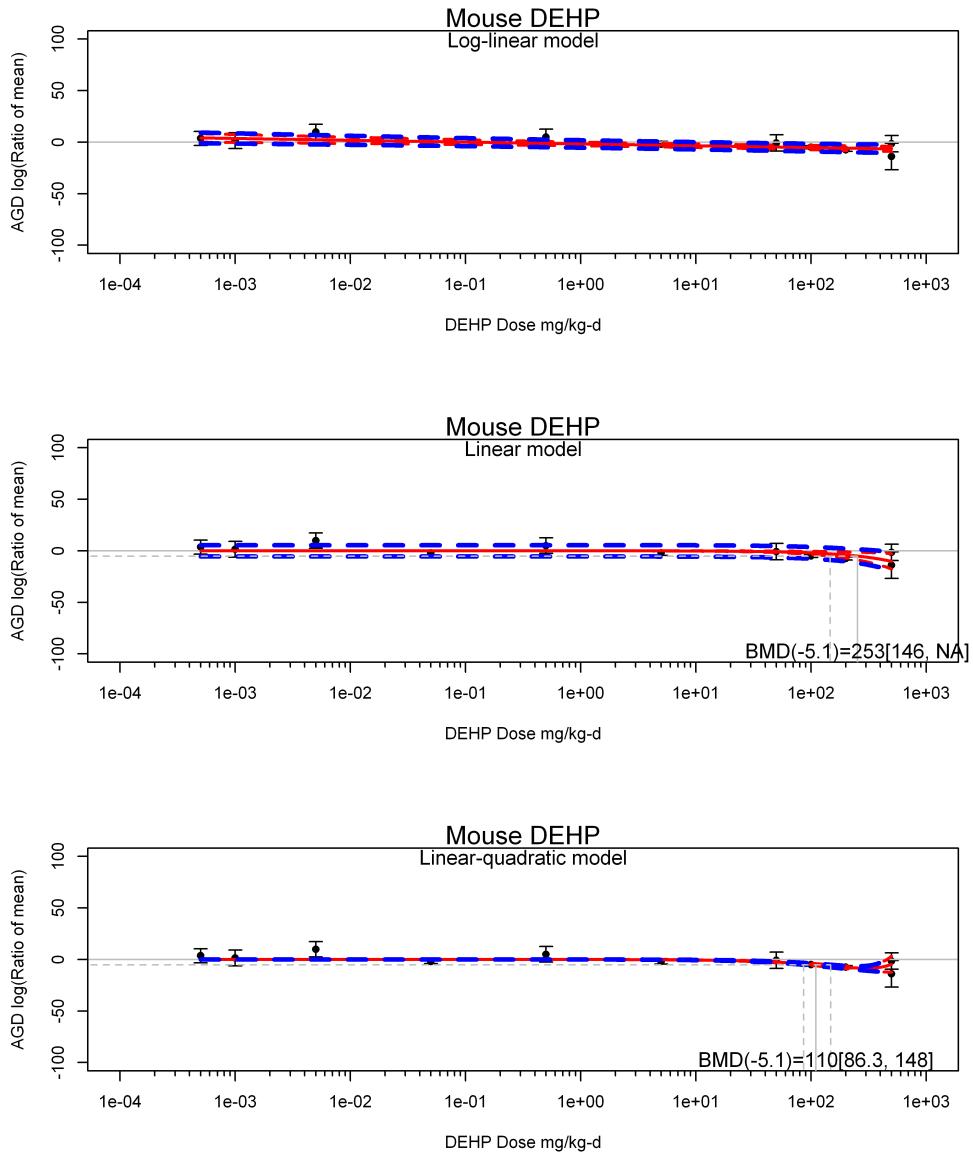


FIGURE C5-4 Benchmark dose estimates from mouse studies of DEHP and AGD.

TABLE C5-5 Benchmark Dose Estimates for DEHP and AGD in Mice

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	253	146	NA
Linear-Quadratic in dose100	-5.1	110	86	148

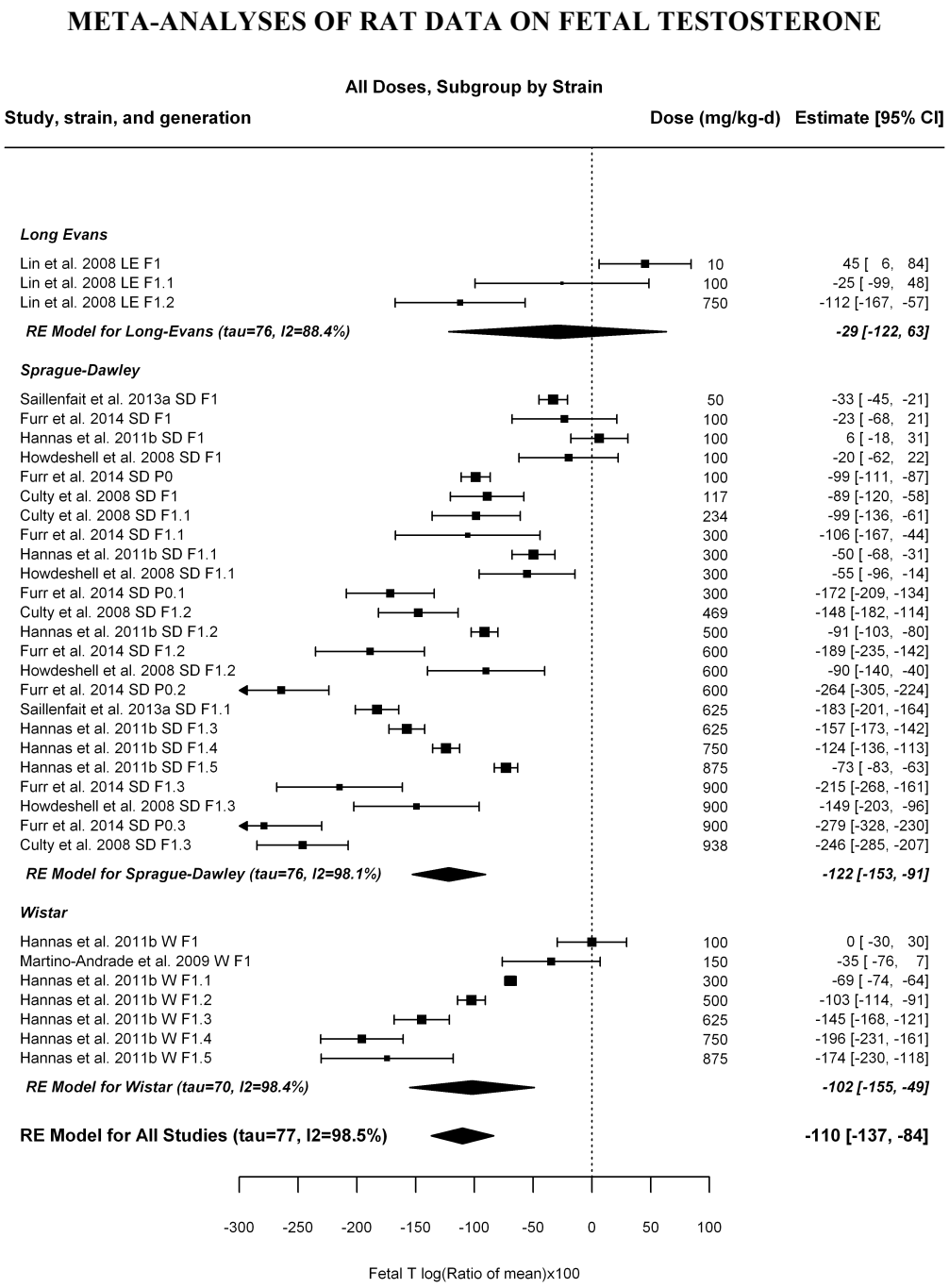


FIGURE C5-5 Results of meta-analyses of studies on DEHP and fetal testosterone in different strains of rat using the random effects model.

TABLE C5-6 Subgrouping Analyses of Rat Studies on DEHP and Fetal Testosterone

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Long Evans									
Rat DEHP LE Overall	intrcpt	-29.3	-121.9	63.3	0.535	76.5	88.4	0.000	39.2
Rat DEHP LE Trend in log10 dose	log10(dose)	-83.0	-118.6	-47.4	0.000	0.0	0.0	0.751	39.0
Rat DEHP LE Linear in dose100	dose100	-15.0	-26.0	-3.9	0.008	31.8	62.0	0.061	36.3*
Sprague-Dawley									
Rat DEHP SD Overall	intrcpt	-121.8	-153.1	-90.5	0.000	76.0	98.1	0.000	270.6
Rat DEHP SD Trend in log10 dose	log10(dose)	-141.8	-200.9	-82.6	0.000	53.7	95.9	0.000	247.2
Rat DEHP SD Linear-Quadratic in dose100	dose100	-38.1	-54.4	-21.8	0.000	53.3	95.8	0.000	246.9*
	I(dose100^2)	1.9	-0.2	4.0	0.075	53.3	95.8	0.000	246.9*
Wistar									
Rat DEHP W Overall	intrcpt	-102.0	-155.2	-48.9	0.000	69.8	98.4	0.000	76.4
Rat DEHP W Trend in log10 dose	log10(dose)	-191.7	-246.1	-137.3	0.000	17.2	77.4	0.010	75.1
Rat DEHP W Linear in dose100	dose100	-22.3	-23.9	-20.7	0.000	3.8	21.2	0.203	57.3*

*Indicates the lowest AICc for each strain.

TABLE C5-7 Overall Analyses and Sensitivity Analyses of Rat Studies of DEHP and Fetal Testosterone Without Subgrouping

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-110.14	-136.73	-83.54	0.000	76.76	98.49	0.000	386.87
Trend in log10(dose)	log10(dose)	-132.83	-171.03	-94.63	0.000	47.74	96.06	0.000	349.39
Linear in dose100	dose100	-23.01	-26.24	-19.77	0.000	48.52	96.55	0.000	358.32
Linear-Quadratic in dose100	dose100	-34.23	-47.02	-21.44	0.000	46.72	95.49	0.000	348.01*
	I(dose100^2)	1.53	-0.16	3.21	0.076				
Sensitivity Analyses									
Overall minus Culty et al. 2008	intrcpt	-105.46	-134.09	-76.82	0.000	77.66	98.64	0.000	341.26
Overall minus Howdeshell et al. 2008	intrcpt	-114.34	-143.48	-85.21	0.000	79.30	98.72	0.000	342.30
Overall minus Lin et al. 2008	intrcpt	-117.28	-144.05	-90.51	0.000	73.90	98.49	0.000	349.80
Overall minus Saillenfait et al. 2013a	intrcpt	-110.31	-137.82	-82.80	0.000	76.92	98.40	0.000	363.95
Overall minus Hannas et al. 2011b	intrcpt	-117.57	-154.20	-80.93	0.000	84.83	96.74	0.000	252.17
Overall minus Martino-Andrade et al. 2009	intrcpt	-112.40	-139.42	-85.37	0.000	76.87	98.52	0.000	375.37
Overall minus Furr et al. 2014	intrcpt	-92.95	-119.61	-66.28	0.000	67.05	98.30	0.000	287.66
Highest Doses-Overall	intrcpt	-162.10	-214.75	-109.45	0.000	77.37	96.29	0.000	99.26
Highest Doses-Trend in log10(dose)	log10(dose)	-195.38	-380.18	-10.58	0.038	64.59	94.51	0.000	92.78
Highest Doses-Linear in dose100	dose100	-21.03	-26.26	-15.79	0.000	60.17	93.59	0.000	95.30
Highest Doses-Linear-Quadratic in dose100	dose100	-21.87	-73.27	29.53	0.404	64.57	92.67	0.000	92.71
	I(dose100^2)	0.10	-5.83	6.02	0.975				

*Indicate the lowest AICc.

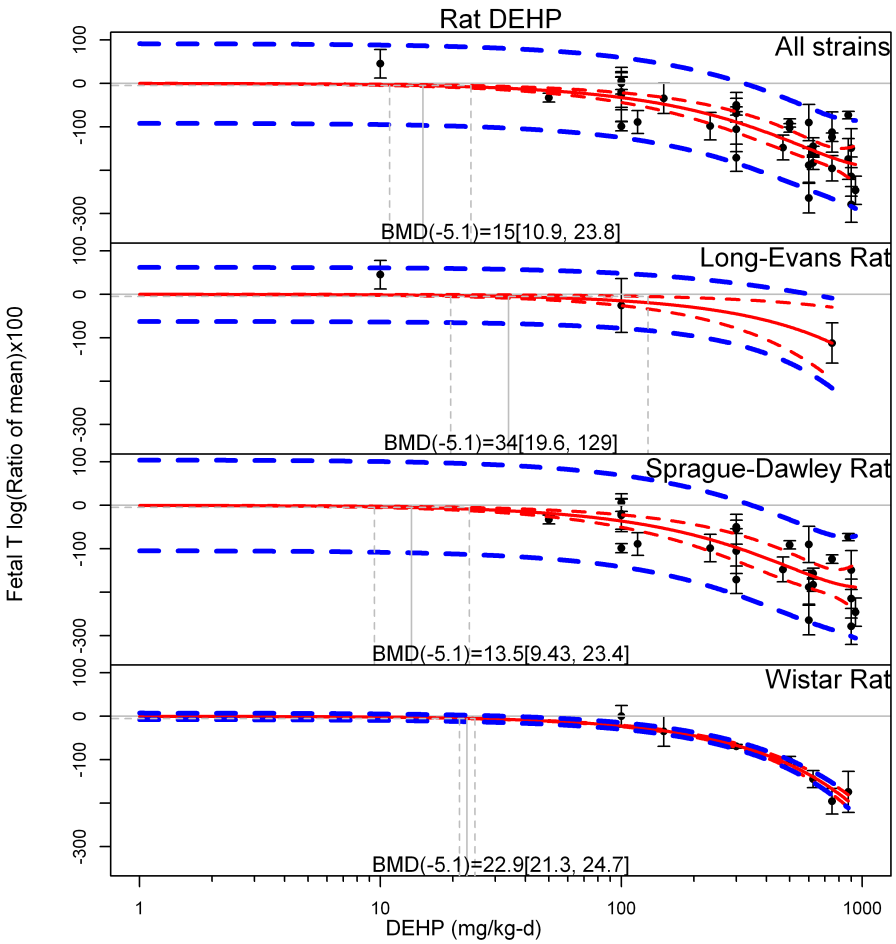


FIGURE C5-6 Benchmark dose estimates from rat studies of DEHP and fetal testosterone (effect size of 5%).

TABLE C5-8 Benchmark Dose Estimates for DEHP and Fetal Testosterone (Effect Size of 5%) in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
All strains Linear in dose100	-5.1	22	20	26
All strains Linear-Quadratic in dose100	-5.1	15	11	24
LE Linear in dose100	-5.1	34	20	129
SD Linear-Quadratic in dose100	-5.1	13	9	23
W Linear in dose100	-5.1	23	21	25

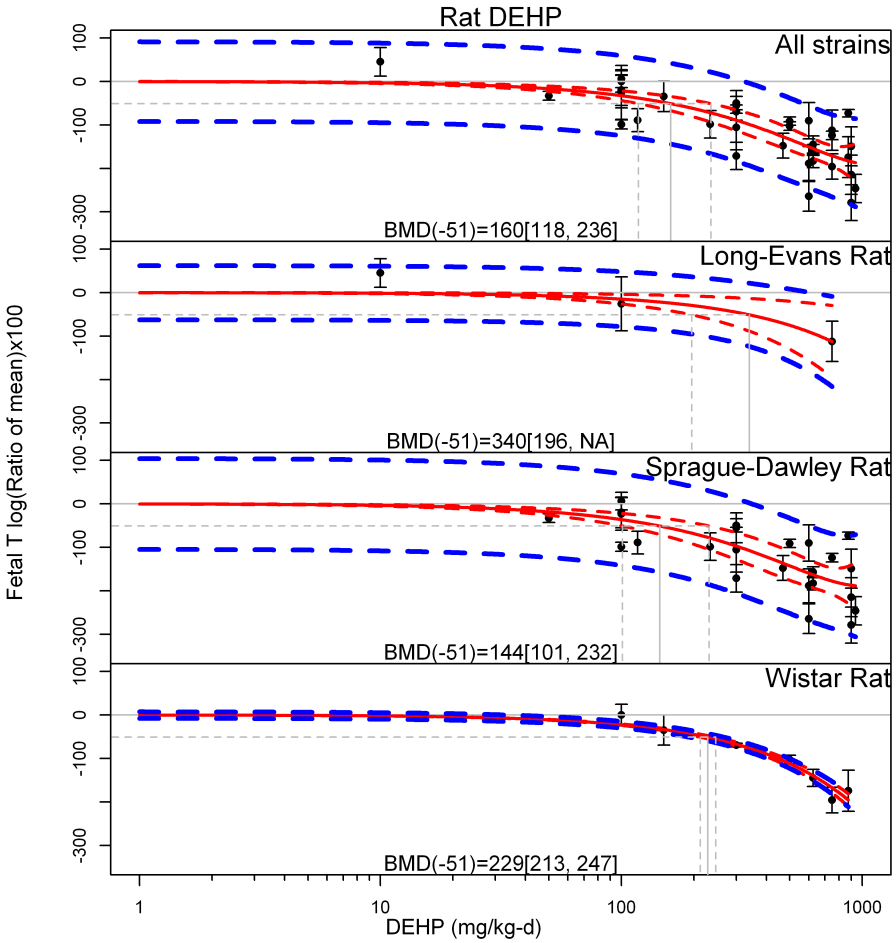


FIGURE C5-7 Benchmark dose estimates from rat studies of DEHP and fetal testosterone (effect size of 40%).

TABLE C5-9 Benchmark Dose Estimates for DEHP and Fetal Testosterone (Effect Size of 40%) in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
All strains Linear in dose100	-51	222	195	258
All strains Linear-Quadratic in dose100	-51	161	118	236
LE Linear in dose100	-51	340	196	NA
SD Linear-Quadratic in dose100	-51	144	101	232
W Linear in dose100	-51	229	213	247

SECTION C-6

Meta-Analyses of Studies of Other Phthalates and AGD or Fetal Testosterone

SUPPLEMENTAL INFORMATION ABOUT METHODS

The conversion from log transformed ratio of means (ROM) to a percent change is as follows:

The effect sizes reported are

$$y = 100 \times ROM = 100 \times \ln(\text{treated response/control response}).$$

Therefore:

$$\text{Percent change} = 100 \times (\text{treated} - \text{control})/\text{control} = 100 \times (\text{treated/control} - 1)$$

$$= 100 \times (\exp(y/100) - 1)$$

BENZYLBUTYL PHTHALATE (BzBP)

For anogenital distance (AGD), there was a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose. There was substantial, statistically significant heterogeneity in all cases (I^2 >75%). The statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. The linear-quadratic model had the lowest AICc (Akaike information criterion corrected for small sample sizes), and benchmark dose estimates from this model were 252 [164, 377] for a 5% change (BMR = -5.1).

For fetal testosterone, there was also a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose, with an overall effect that is large in magnitude (>50% change). There was substantial, statistically significant heterogeneity in all cases (I^2 >85%). There were too few studies for sensitivity analyses. The linear-quadratic model had the lowest AICc, and benchmark dose estimates from this model were 23 mg/kg-day [95% CI: 13, 74] for a 5% change (BMR = -5.1) and 230 mg/kg-day [140, 390] for a 40% change (BMR = 51).

Although there was substantial heterogeneity, standard deviation of the random effect (tau) was less than the estimated size of the effect at higher doses. Therefore, the heterogeneity does not affect the conclusion that BzBP exposure affects both AGD and fetal testosterone in the rat.

DIBUTYL PHTHALATE (DBP)

For AGD there was a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose. There was substantial, statistically significant heterogeneity in all cases (I^2 >75%). The statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. The linear-quadratic model had the lowest AICc, and benchmark dose estimate from this model wsd 153 mg/kg-day [95% CI: 115, 216] for a 5% change (BMR = -5.1).

TABLE C6-1 Overall Analyses and Sensitivity Analyses of Rat Studies of BzBP and AGD

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-5.34	-8.35	-2.33	0.001	5.78	94.87	0.000	102.07
Trend in log10(dose)	log10(dose)	-4.68	-7.09	-2.27	0.000	3.91	88.95	0.000	90.43
Linear in dose100	dose100	-2.07	-2.50	-1.63	0.000	2.36	76.64	0.000	82.92
Linear-Quadratic in dose100	dose100	-2.01	-3.45	-0.56	0.007	2.50	77.70	0.000	82.14*
	I(dose100^2)	-0.01	-0.24	0.22	0.927				
Sensitivity Analyses									
Overall minus Ashby et al. 1997	intrcpt	-6.00	-8.92	-3.08	0.000	5.36	93.79	0.000	94.06
Overall minus Aso et al. 2005	intrcpt	-5.41	-9.65	-1.16	0.012	6.68	97.37	0.000	66.33
Overall minus Tyl et al. 2004	intrcpt	-3.72	-6.82	-0.63	0.018	4.47	90.78	0.000	60.65
Overall minus Nagao et al. 2000	intrcpt	-5.72	-9.38	-2.06	0.002	6.32	93.59	0.000	84.97
Highest Doses-Overall	intrcpt	-8.57	-15.41	-1.72	0.014	8.24	95.86	0.000	45.55
Highest Doses-Linear in dose100	dose100	-2.02	-2.59	-1.45	0.000	3.15	79.09	0.000	38.03
Highest Doses-Trend in log10(dose)	log10(dose)	-4.54	-7.89	-1.19	0.008	5.27	86.99	0.000	55.54
Highest Doses-Linear-Quadratic in dose100	dose100	-0.82	-3.61	1.98	0.566	3.17	78.74	0.001	52.64
Highest Doses-Linear-Quadratic in dose100	I(dose100^2)	-0.18	-0.60	0.23	0.388	3.17	78.74	0.001	52.64

*Indicates the lowest AICc.

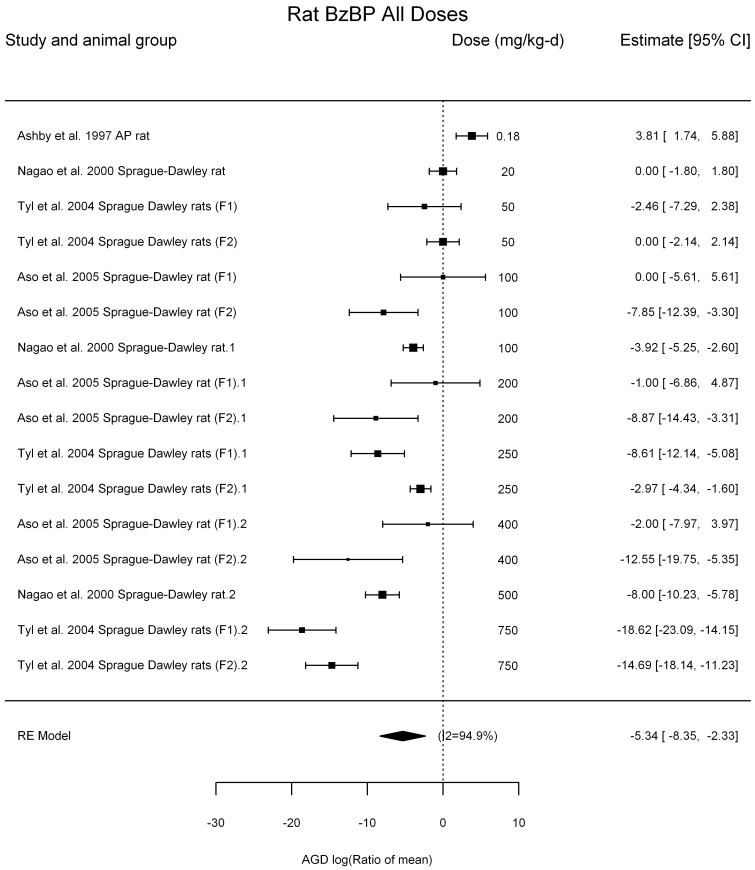


FIGURE C6-1 Meta-analyses of studies of BzBP and AGD in rats.

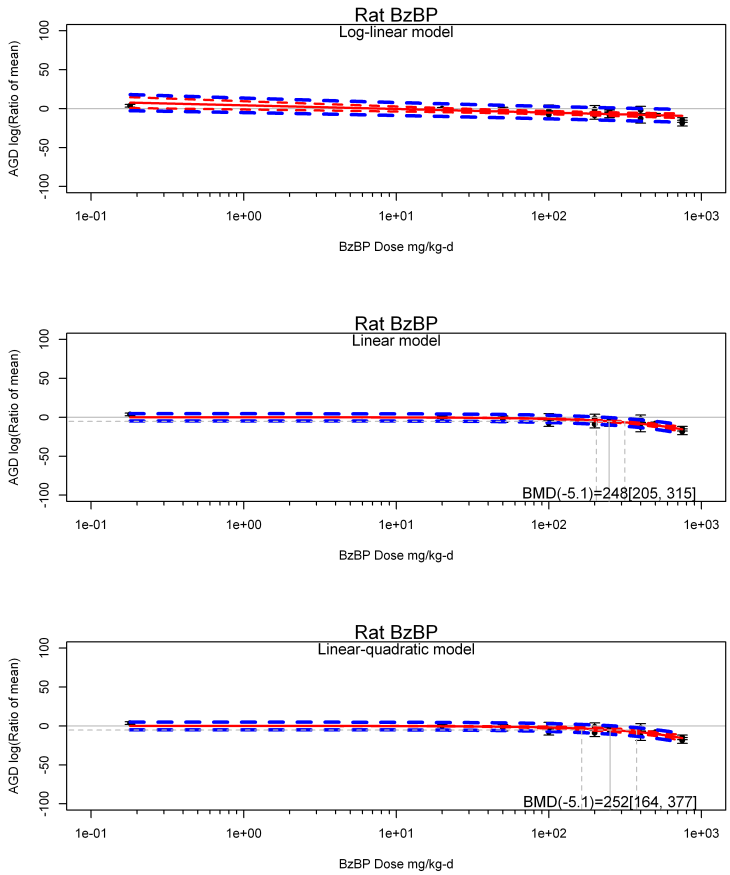


FIGURE C6-2 Benchmark dose estimates from rat studies of BzBP and AGD.

TABLE C6-2 Benchmark Dose Estimates for BzBP and AGD in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	248	205	315
Linear-Quadratic in dose100	-5.1	252	164	377

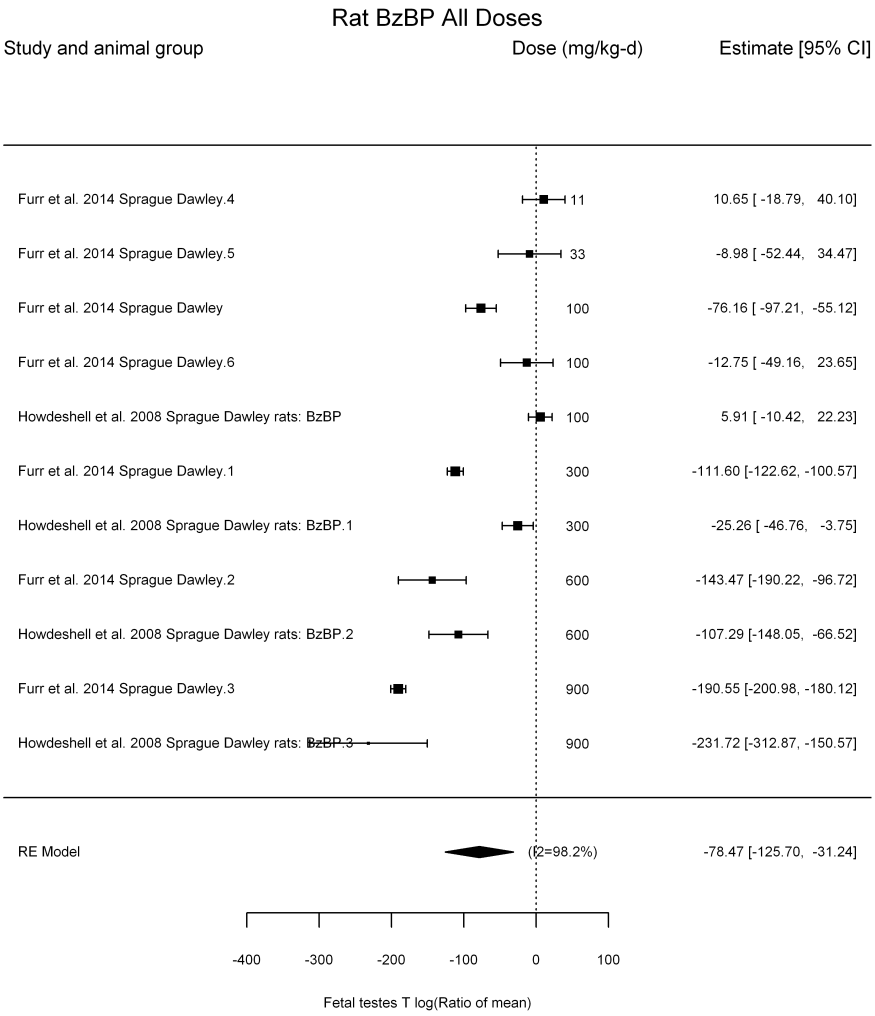


FIGURE C6-3 Meta-analyses of studies of BzBP and fetal testosterone in rats.

TABLE C6-3 Overall Analyses of Rat Studies of BzBP and Fetal Testosterone*

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Overall	intrcpt	-78.47	-125.70	-31.24	0.001	77.72	98.17	0.000	122.09
Trend in log10(dose)	log10(dose)	-106.74	-154.77	-58.71	0.000	43.83	93.93	0.000	105.93
Linear in dose100	dose100	-22.12	-26.60	-17.64	0.000	29.98	87.79	0.000	103.86
Linear-Quadratic in dose100	dose100	-22.52	-39.59	-5.45	0.010	31.76	86.02	0.000	100.00**
	I(dose100^2)	0.05	-2.14	2.24	0.964				

*Too few studies for sensitivity analyses.
**Indicates the lowest AICc.

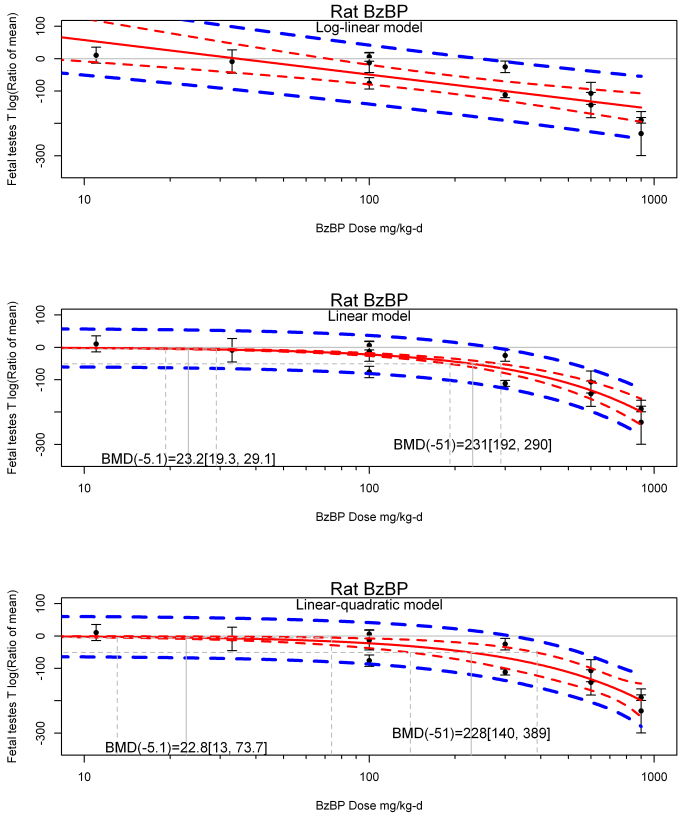


FIGURE C6-4 Benchmark dose estimates from rat studies of BzBP and fetal testosterone.

TABLE C6-4 Benchmark Dose Estimates for BzBP and Fetal Testosterone in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	23	19	29
Linear in dose100	-51.1	231	192	290
Linear-Quadratic in dose100	-5.1	23	13	74
Linear-Quadratic in dose100	-51.1	228	140	389

For fetal testosterone there was also a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose, with an overall effect that is large in magnitude (>50% change). There was substantial, statistically significant heterogeneity in all cases (I^2 >80%). The statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. The linear-quadratic model had the lowest AICc, and benchmark dose estimates from this model were 12 mg/kg-day [95% CI: 8, 22] for a 5% change (BMR = -5.1) and 130 mg/kg-day [95% CI: 85, 210] for a 40% change (BMR = 51).

Although there was substantial heterogeneity, standard deviation of the random effect (tau) was less than the estimated size of the effect at higher doses. Therefore, the heterogeneity does not affect the conclusion that DBP exposure affects both AGD and fetal testosterone in the rat.

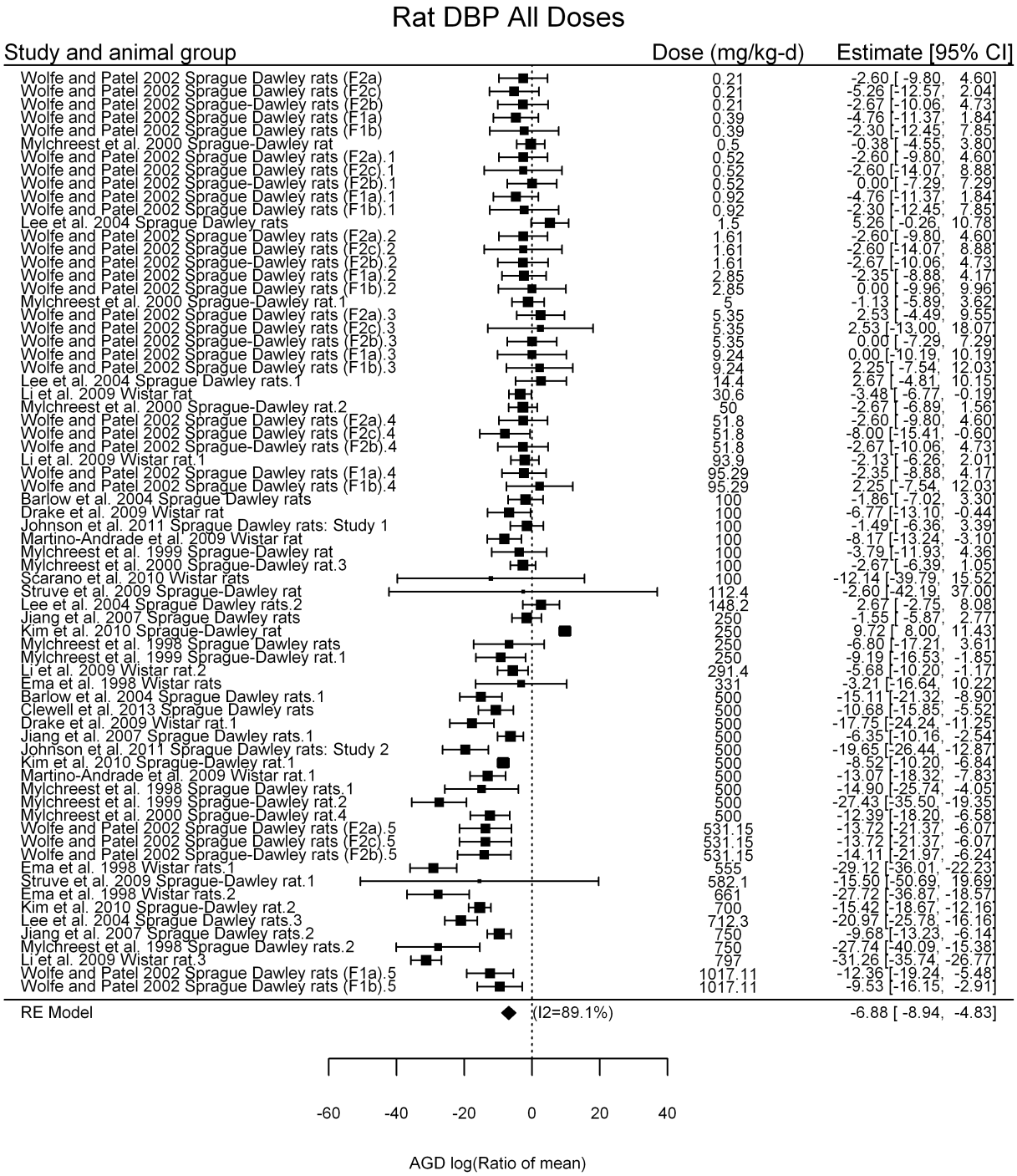


FIGURE C6-5 Meta-analyses of studies of DBP and AGD in rats.

TABLE C6-5 Overall Analyses and Sensitivity Analyses of Rat Studies of DBP and AGD

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-6.88	-8.94	-4.83	0.000	7.84	89.12	0.000	500.28
Trend in log10(dose)	log10(dose)	-4.14	-5.63	-2.65	0.000	6.38	84.31	0.000	471.15
Linear in dose100	dose100	-2.42	-2.80	-2.04	0.000	4.95	76.58	0.000	449.24
Linear-Quadratic in dose100	dose100	-3.64	-4.85	-2.42	0.000	4.90	75.57	0.000	441.75*
	I(dose100^2)	0.18	0.01	0.35	0.039				
Sensitivity Analyses									
Overall minus Struve et al. 2009	intrcpt	-6.87	-8.94	-4.80	0.000	7.86	89.45	0.000	484.30
Overall minus Barlow et al. 2004	intrcpt	-6.84	-8.94	-4.73	0.000	7.89	89.27	0.000	486.80
Overall minus Li et al. 2009	intrcpt	-6.61	-8.64	-4.59	0.000	7.35	87.25	0.000	466.34
Overall minus Johnson et al. 2011	intrcpt	-6.78	-8.86	-4.69	0.000	7.79	89.03	0.000	485.44
Overall minus Mylchreest et al. 1998	intrcpt	-6.55	-8.62	-4.48	0.000	7.74	89.20	0.000	476.04
Overall minus Jiang et al. 2007	intrcpt	-6.94	-9.09	-4.78	0.000	8.04	89.17	0.000	481.61
Overall minus Mylchreest et al. 2000	intrcpt	-7.16	-9.37	-4.96	0.000	8.07	89.27	0.000	467.83
Overall minus Mylchreest et al. 1999	intrcpt	-6.59	-8.65	-4.53	0.000	7.66	88.89	0.000	476.03
Overall minus Scarano et al. 2010	intrcpt	-6.86	-8.93	-4.79	0.000	7.86	89.30	0.000	492.79
Overall minus Kim et al. 2010	intrcpt	-7.01	-9.07	-4.95	0.000	7.60	85.07	0.000	476.65
Overall minus Drake et al. 2009	intrcpt	-6.71	-8.81	-4.62	0.000	7.86	89.26	0.000	486.40
Overall minus Lee et al. 2004	intrcpt	-7.15	-9.21	-5.09	0.000	7.54	88.37	0.000	468.14
Overall minus Martino-Andrade et al. 2009	intrcpt	-6.76	-8.88	-4.64	0.000	7.94	89.35	0.000	487.55
Overall minus Wolfe and Patel 2002	intrcpt	-9.17	-12.36	-5.98	0.000	9.50	94.35	0.000	297.19
Overall minus Ema et al. 1998	intrcpt	-6.29	-8.23	-4.34	0.000	7.16	87.54	0.000	468.32
Overall minus Clewell et al. 2013	intrcpt	-6.82	-8.91	-4.73	0.000	7.90	89.25	0.000	494.00
Highest Doses-Overall	intrcpt	-16.07	-19.41	-12.74	0.000	6.71	83.07	0.000	143.78
Highest Doses-Linear in dose100	dose100	-2.49	-3.03	-1.95	0.000	7.00	84.14	0.000	145.84
Highest Doses-Trend in log10(dose)	log10(dose)	-14.44	-28.07	-0.81	0.038	5.99	79.18	0.000	136.70
Highest Doses-Linear-Quadratic in dose100	dose100	-5.20	-6.93	-3.48	0.000	5.48	76.59	0.000	134.34
	I(dose100^2)	0.37	0.14	0.60	0.001				

*Indicates the lowest AICc.

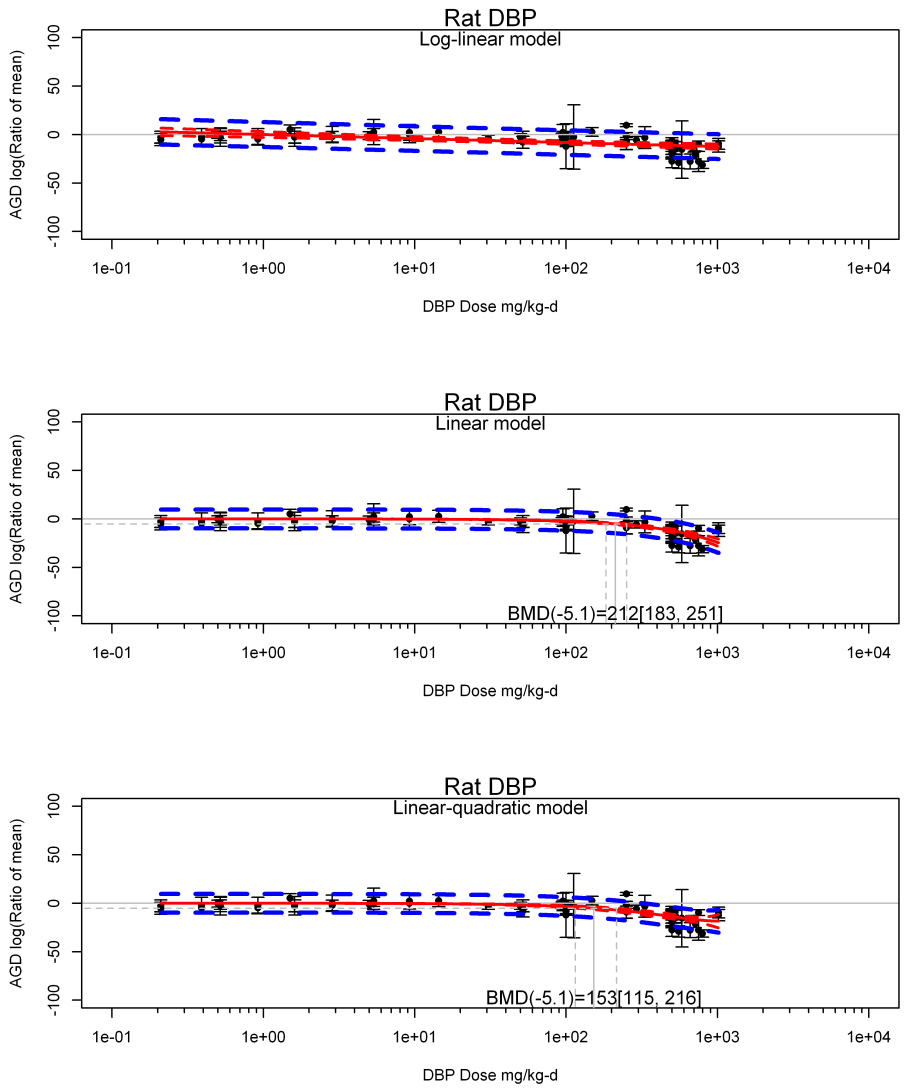


FIGURE C6-6 Benchmark dose estimates from rat studies of DBP and AGD.

TABLE C6-6 Benchmark Dose Estimates for DBP and AGD in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	212	183	251
Linear-Quadratic in dose100	-5.1	153	115	216

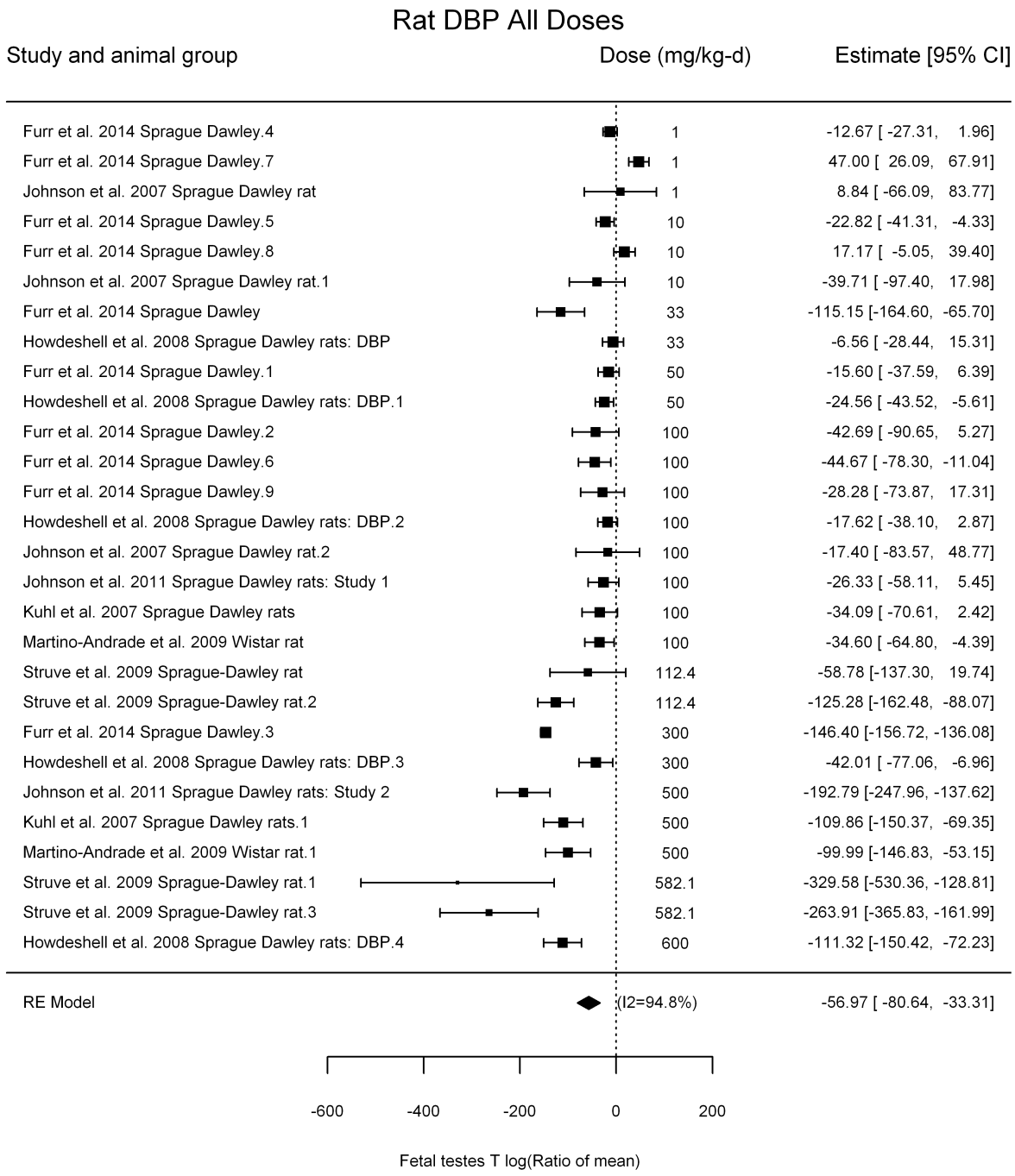


FIGURE C6-7 Meta-analyses of studies of DBP and fetal testosterone in rats.

TABLE C6-7 Overall Analyses and Sensitivity Analyses of Rat Studies of DBP and Fetal Testosterone

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-56.97	-80.64	-33.31	0.000	59.25	94.78	0.000	311.44
Trend in log10(dose)	log10(dose)	-53.72	-74.64	-32.79	0.000	39.50	88.01	0.000	285.61
Linear in dose100	dose100	-29.43	-35.83	-23.04	0.000	35.25	86.12	0.000	285.72
Linear-Quadratic in dose100	dose100	-44.98	-66.52	-23.43	0.000	32.90	83.99	0.000	277.00*
	I(dose100^2)	3.29	-1.05	7.63	0.137				
Sensitivity Analyses									
Overall minus Struve et al. 2009	intrcpt	-45.23	-67.11	-23.35	0.000	51.20	93.92	0.000	254.23
Overall minus Howdeshell et al. 2008	intrcpt	-62.19	-90.88	-33.50	0.000	65.13	95.26	0.000	258.67
Overall minus Johnson et al. 2007	intrcpt	-61.63	-87.60	-35.66	0.000	62.18	95.65	0.000	279.22
Overall minus Johnson et al. 2011	intrcpt	-52.63	-75.58	-29.69	0.000	54.84	94.16	0.000	286.41
Overall minus Kuhl et al. 2007	intrcpt	-56.08	-81.36	-30.80	0.000	61.09	95.24	0.000	290.39
Overall minus Martino-Andrade et al. 2009	intrcpt	-56.61	-82.07	-31.15	0.000	61.58	95.28	0.000	290.63
Overall minus Furr et al. 2014	intrcpt	-70.79	-101.69	-39.89	0.000	60.74	91.61	0.000	201.05
Highest Doses-Overall	intrcpt	-116.72	-164.82	-68.62	0.000	71.04	94.06	0.000	112.13
Highest Doses-Trend in log10(dose)	log10(dose)	-160.89	-246.19	-75.60	0.000	37.54	81.20	0.000	99.44
Highest Doses-Linear in dose100	dose100	-29.77	-37.50	-22.05	0.000	42.45	86.55	0.000	104.16
Highest Doses-Linear-Quadratic in dose100	dose100	-49.92	-86.82	-13.02	0.008	37.92	77.06	0.000	99.67
	I(dose100^2)	3.98	-3.08	11.04	0.269				

*Indicates the lowest AICc.

TABLE C6-8 Benchmark Dose Estimates for DBP and Fetal Testosterone in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	17	14	22
Linear in dose100	-51.1	174	143	222
Linear-Quadratic in dose100	-5.1	12	8	22
Linear-Quadratic in dose100	-51.1	125	85	205

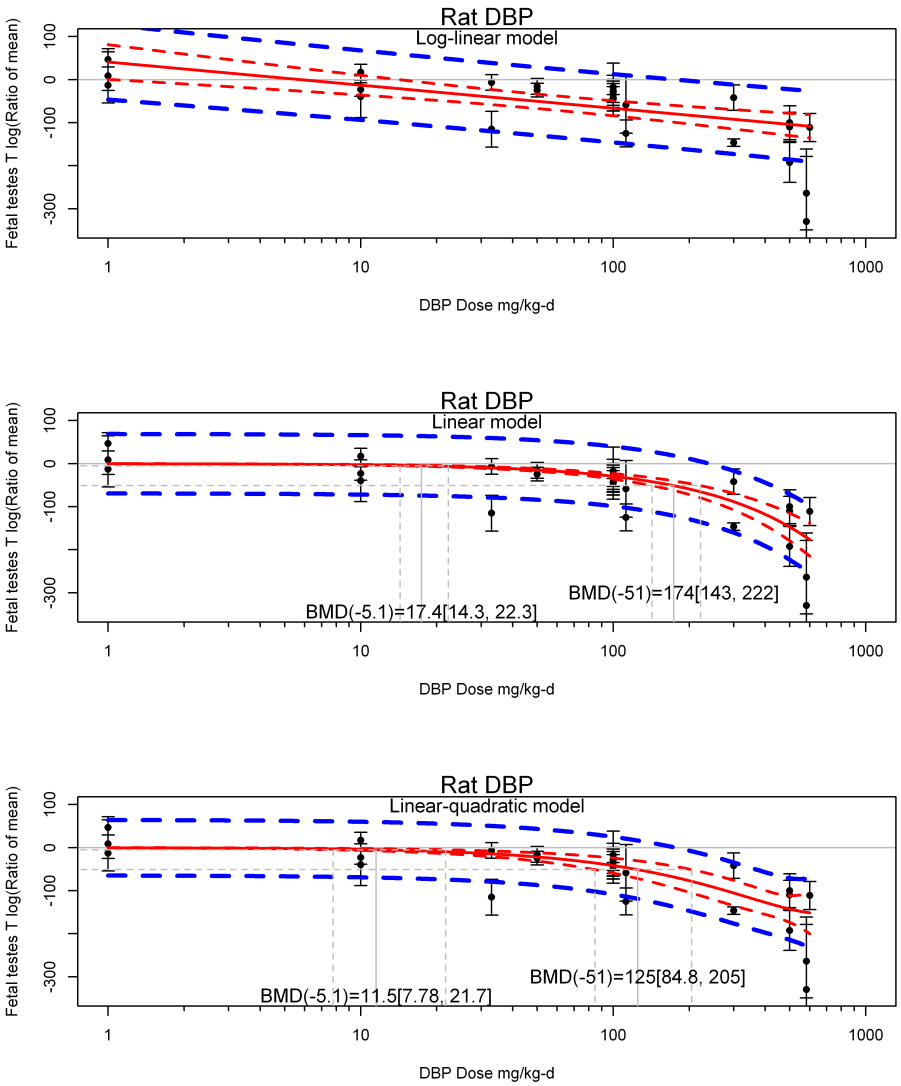


FIGURE C6-8 Benchmark dose estimates from rat studies of DBP and fetal testosterone.

DIPENTYL PHTHALATE (DPP)

For fetal testosterone, there was also a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose, with an overall effect that is large in magnitude (>50% change). There was substantial, statistically significant heterogeneity in all cases (I^2 >90%). The statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. The linear-quadratic model had the lowest AICc, and benchmark dose estimates from this model were 5.6 [95% CI: 4.8, 6.4] for a 5% change (BMR = -5.1) and 58 [95% CI: 50, 70] for a 40% change (BMR = 51).

Although there was substantial heterogeneity, standard deviation of the random effect (tau) was less than the estimated size of the effect at higher doses. Therefore, the heterogeneity does not affect the conclusion that DPP exposure affects fetal testosterone in the rat.

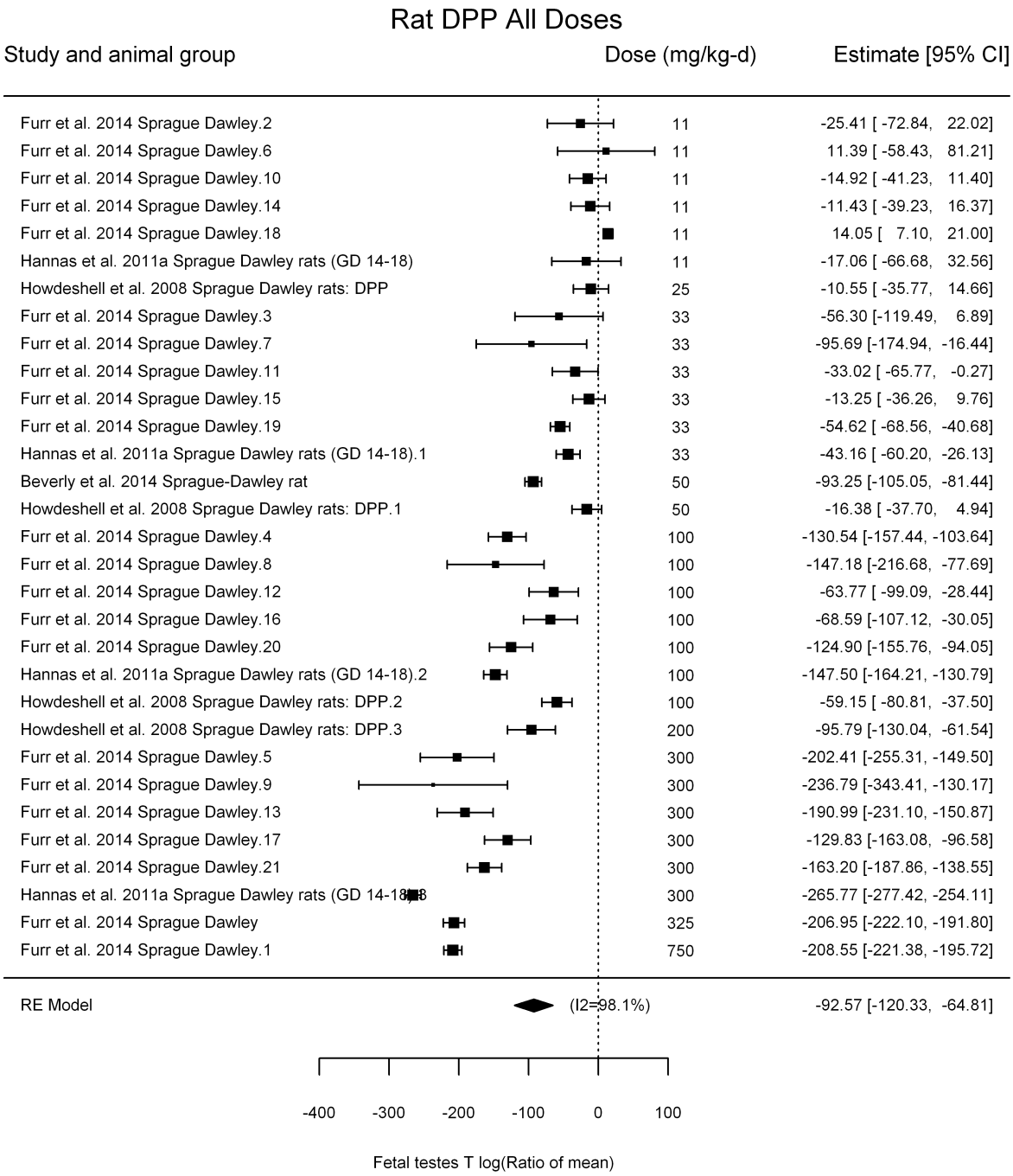


FIGURE C6-9 Meta-analyses of studies of DPP and fetal testosterone in rats.

TABLE C6-9 Overall Analyses and Sensitivity Analyses of Rat Studies of DPP and Fetal Testosterone

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-92.57	-120.33	-64.81	0.000	76.25	98.14	0.000	351.76
Trend in log10(dose)	log10(dose)	-127.64	-152.92	-102.36	0.000	34.29	90.42	0.000	300.49
Linear in dose100	dose100	-50.24	-60.17	-40.30	0.000	56.12	96.75	0.000	334.20
Linear-Quadratic in dose100	dose100	-93.99	-107.96	-80.02	0.000	32.57	90.61	0.000	298.59*
	I(dose100^2)	8.93	6.44	11.42	0.000				
Sensitivity Analyses									
Overall minus Howdeshell et al. 2008	intrcpt	-99.93	-130.63	-69.23	0.000	78.63	98.31	0.000	307.12
Overall minus Beverly et al. 2014	intrcpt	-92.57	-121.32	-63.81	0.000	77.71	98.07	0.000	341.24
Overall minus Hannas et al. 2011a	intrcpt	-88.02	-115.86	-60.18	0.000	70.84	97.62	0.000	302.42
Overall minus Furr et al. 2014	intrcpt	-84.01	-138.21	-29.81	0.002	81.93	98.71	0.000	99.72
Highest Doses-Overall	intrcpt	-173.32	-214.05	-132.59	0.000	58.87	96.97	0.000	95.02
Highest Doses-Trend in log10(dose)	log10(dose)	-117.01	-227.86	-6.15	0.039	48.75	94.22	0.000	89.25
Highest Doses-Linear in dose100	dose100	-45.89	-61.35	-30.43	0.000	82.45	98.43	0.000	100.18
Highest Doses-Linear-Quadratic in dose100	dose100	-87.62	-111.52	-63.72	0.000	48.60	94.72	0.000	89.05
	I(dose100^2)	7.96	3.88	12.05	0.000				

*Indicates the lowest AICc.

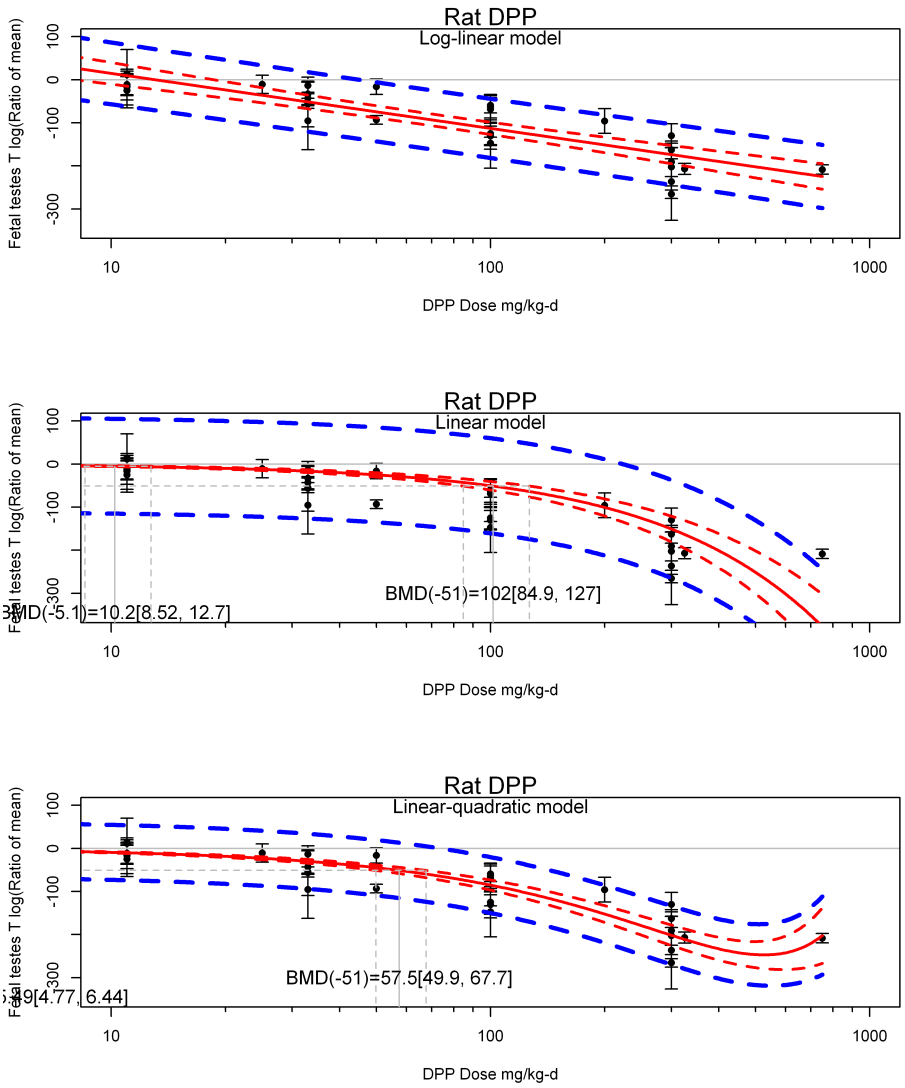


FIGURE C6-10 Benchmark dose estimates from rat studies of DPP and fetal testosterone.

TABLE C6-10 Benchmark Dose Estimates for DPP and Fetal Testosterone in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	10	9	13
Linear in dose100	-51.1	102	85	127
Linear-Quadratic in dose100	-5.1	5.6	4.8	6.4
Linear-Quadratic in dose100	-51.1	58	50	68

DIISOBUTYL PHTHALATE (DIBP)

For fetal testosterone, there was also a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose, with an overall effect that is large in magnitude (>50% change). There was substantial, statistically significant heterogeneity in all cases (I^2 >60%). There were too few studies to conduct sensitivity analyses. The linear model had the lowest AICc, and a benchmark dose estimate 270 [225,

340] for a 40% change (BMR = 51). (The 5% change was well below the range of the data but will be 10-times lower because a linear model was used.)

Although there was substantial heterogeneity, standard deviation of the random effect (tau) was less than the estimated size of the effect at higher doses. Therefore, the heterogeneity does not affect the conclusion that DIBP exposure affects fetal testosterone in the rat.

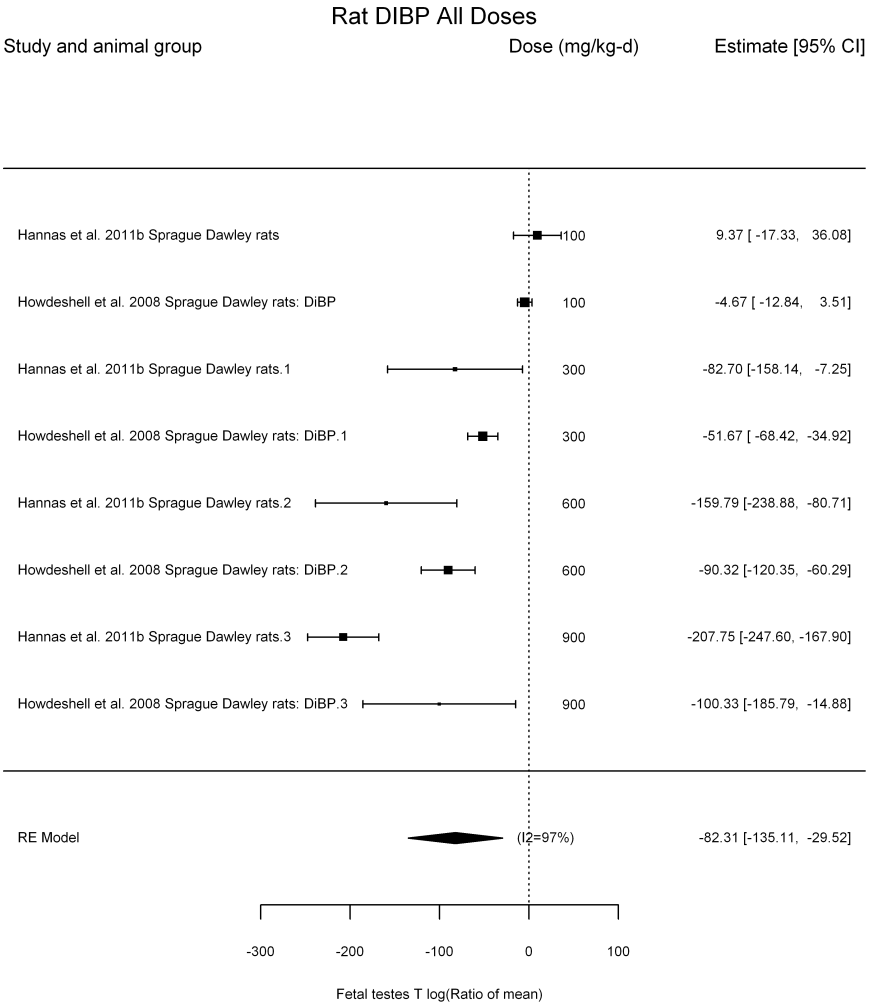


FIGURE C6-11 Meta-analyses of studies of DIBP and fetal testosterone in rats.

TABLE C6-11 Overall Analyses of Rat Studies of DIBP and Fetal Testosterone*

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Overall	intrcpt	-82.31	-135.11	-29.52	0.002	71.76	96.96	0.000	87.28
Trend in log10(dose)	log10(dose)	-169.23	-234.13	-104.33	0.000	28.14	77.83	0.001	78.52
Linear in dose100	dose100	-18.84	-22.73	-14.94	0.000	18.64	78.78	0.001	75.51**
Linear-Quadratic in dose100	dose100	-11.61	-22.13	-1.08	0.031	12.22	57.12	0.020	77.04
	I(dose100^2)	-1.00	-2.42	0.42	0.169				

*Too few studies for sensitivity analyses.
**Indicates the lowest AICc.

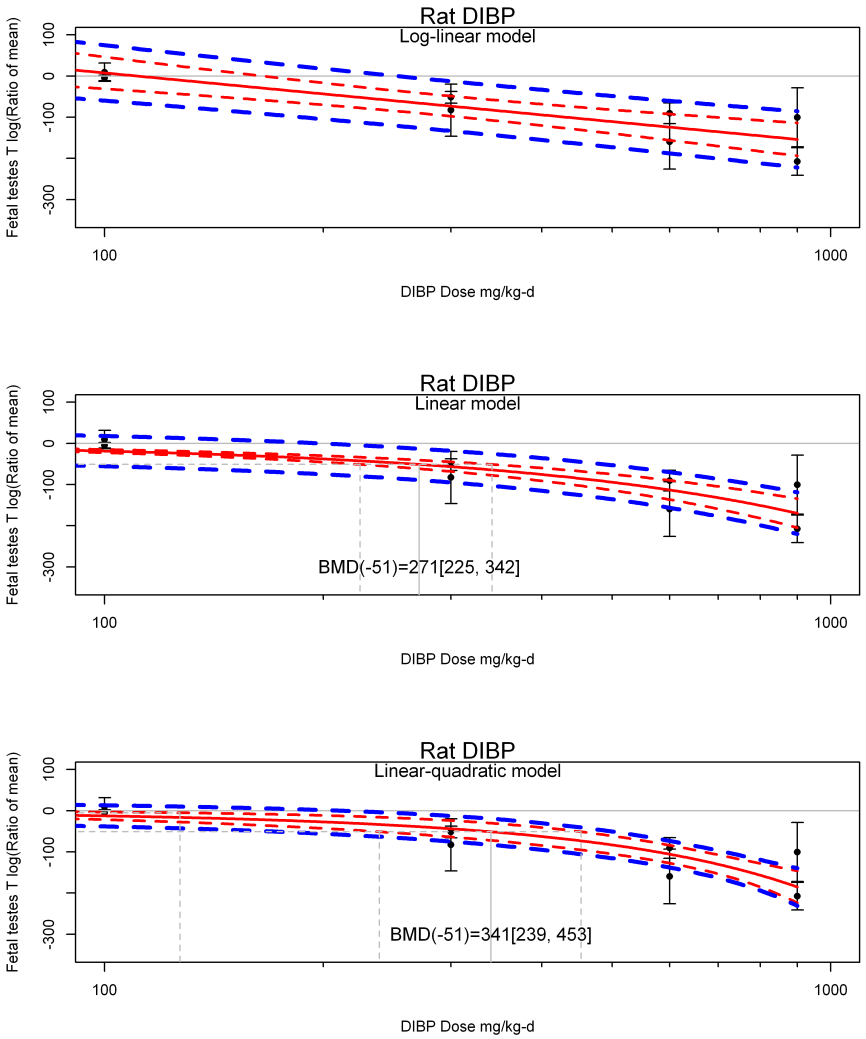


FIGURE C6-12 Benchmark dose estimates from rat studies of DIBP and fetal testosterone.

TABLE C6-12 Benchmark Dose Estimates for DIBP and Fetal Testosterone in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	27	23	34
Linear in dose100	-51.1	271	225	342
Linear-Quadratic in dose100	-5.1	43	23	127
Linear-Quadratic in dose100	-51.1	341	239	453

DIISONONYL PHTHALATE (DINP)

For AGD, there was no statistically significant overall effect, nor were there any statistically significant trends in $\log_{10}$(dose) or dose. There was very little heterogeneity in all cases (I^2 <5%). The lack of statistical significance of these effects was robust to leaving out individual studies and restricting to the highest dose group from each study. The linear model had the lowest AICc, and due to the lack of statistically significant trend, the upper confidence limit on the benchmark was unbounded, and only a lower confidence bound of 684 could be derived for a 5% change (BMR = -5.1). In sum, although a small effect

was observed, the precision of the estimate was not sufficient to rule out chance. Thus, the available studies do not support DINP exposure being associated with decreased AGD.

By contrast, for fetal testosterone, there was a statistically significant overall effect and linear trends in $\log_{10}$(dose) and dose, with an overall effect that is large in magnitude (>50% change). There was substantial heterogeneity in the overall effect (I^2 = 83%), but this was reduced when the effect of dose was included. I^2 = 42% for trend in $\log_{10}$(dose) and I^2 of 22-24% under a linear or linear-quadratic model, neither of which was statistically significant. There were too few studies to conduct sensitivity analyses. The linear-quadratic model had the lowest AICc, and benchmark dose estimates from this model were 76 mg/kg-day [95% CI: 49, 145] for a 5% change (BMR = -5.1) and 701 mg/kg-day [94% CI: 552, 847] for a 40% change (BMR = 51).

Although there was substantial heterogeneity, standard deviation of the random effect (tau) was less than the estimated size of the effect at higher doses. Therefore, the heterogeneity does not affect the conclusion that DINP exposure affects both AGD and fetal testosterone in the rat.

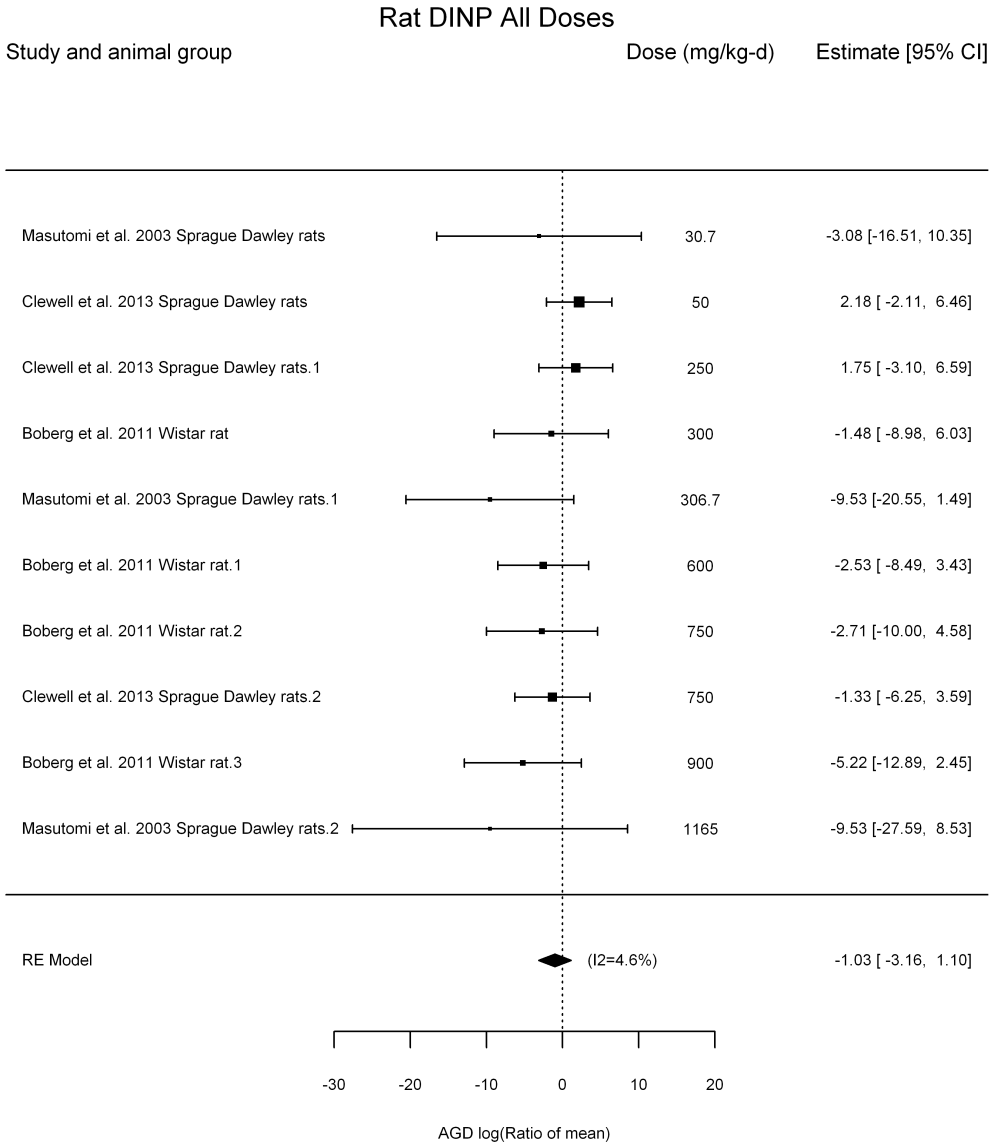


FIGURE C6-13 Meta-analyses of studies of DINP and AGD in rats.

TABLE C6-13 Overall Analyses and Sensitivity Analyses of Rat Studies of DINP and AGD

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	-1.03	-3.16	1.10	0.345	0.75	4.63	0.508	55.61
Trend in log10(dose)	log10(dose)	-3.86	-8.11	0.39	0.075	0.00	0.00	0.748	54.13
Linear in dose100	dose100	-0.36	-0.75	0.02	0.065	0.00	0.00	0.773	52.71*
Linear-Quadratic in dose100	dose100	0.22	-1.33	1.78	0.778	0.00	0.00	0.749	53.03
	I(dose100^2)	-0.08	-0.28	0.12	0.444				
Sensitivity Analyses									
Overall minus Boberg et al. 2011	intrcpt	0.09	-2.45	2.62	0.947	0.00	0.00	0.315	39.75
Overall minus Masutomi et al. 2003	intrcpt	-0.45	-2.58	1.68	0.679	0.00	0.00	0.576	37.15
Overall minus Clewell et al. 2013	intrcpt	-3.67	-6.86	-0.49	0.024	0.00	0.00	0.901	40.31
Highest Doses-Overall	intrcpt	-2.82	-6.85	1.22	0.171	0.00	0.00	0.533	27.40
Highest Doses-Linear in dose100	dose100	-0.38	-0.87	0.12	0.134	0.00	0.00	0.642	26.72
Highest Doses-Trend in log10(dose)	log10(dose)	-45.42	-125.00	34.16	0.263	0.00	0.00	0.930	34.90
Highest Doses-Linear-Quadratic in dose100	dose100	1.09	-2.18	4.37	0.513	0.00	0.00	0.761	34.66
	I(dose100^2)	-0.17	-0.56	0.21	0.373				

*Indicates the lowest AICc.

TABLE C6-14 Benchmark Dose Estimates for DINP and AGD in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	NA	684	NA
Linear-Quadratic in dose100	-5.1	NA	706	NA

TABLE C6-15 Overall Analyses of Rat Studies of DINP and Fetal Testosterone*

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Overall	intrcpt	-63.95	-86.35	-41.55	0.000	31.16	83.28	0.000	118.16
Trend in log10(dose)	log10(dose)	-128.35	-186.46	-70.24	0.000	12.59	42.01	0.076	106.24
Linear in dose100	dose100	-7.56	-8.69	-6.43	0.000	7.21	21.84	0.215	107.77
Linear-Quadratic in dose100	dose100	-6.74	-10.51	-2.96	0.000	8.04	23.81	0.182	104.59**
	I(dose100^2)	-0.08	-0.42	0.26	0.648				

*Too few studies for sensitivity analyses.
**Indicates the lowest AICc.

TABLE C6-16 Benchmark Dose Estimates for DINP and Fetal Testosterone in Rats

Analysis	BMR	BMD	CI, Lower Bound	CI, Upper Bound
Linear in dose100	-5.1	68	59	80
Linear in dose100	-51.1	676	588	795
Linear-Quadratic in dose100	-5.1	76	49	145
Linear-Quadratic in dose100	-51.1	701	552	847

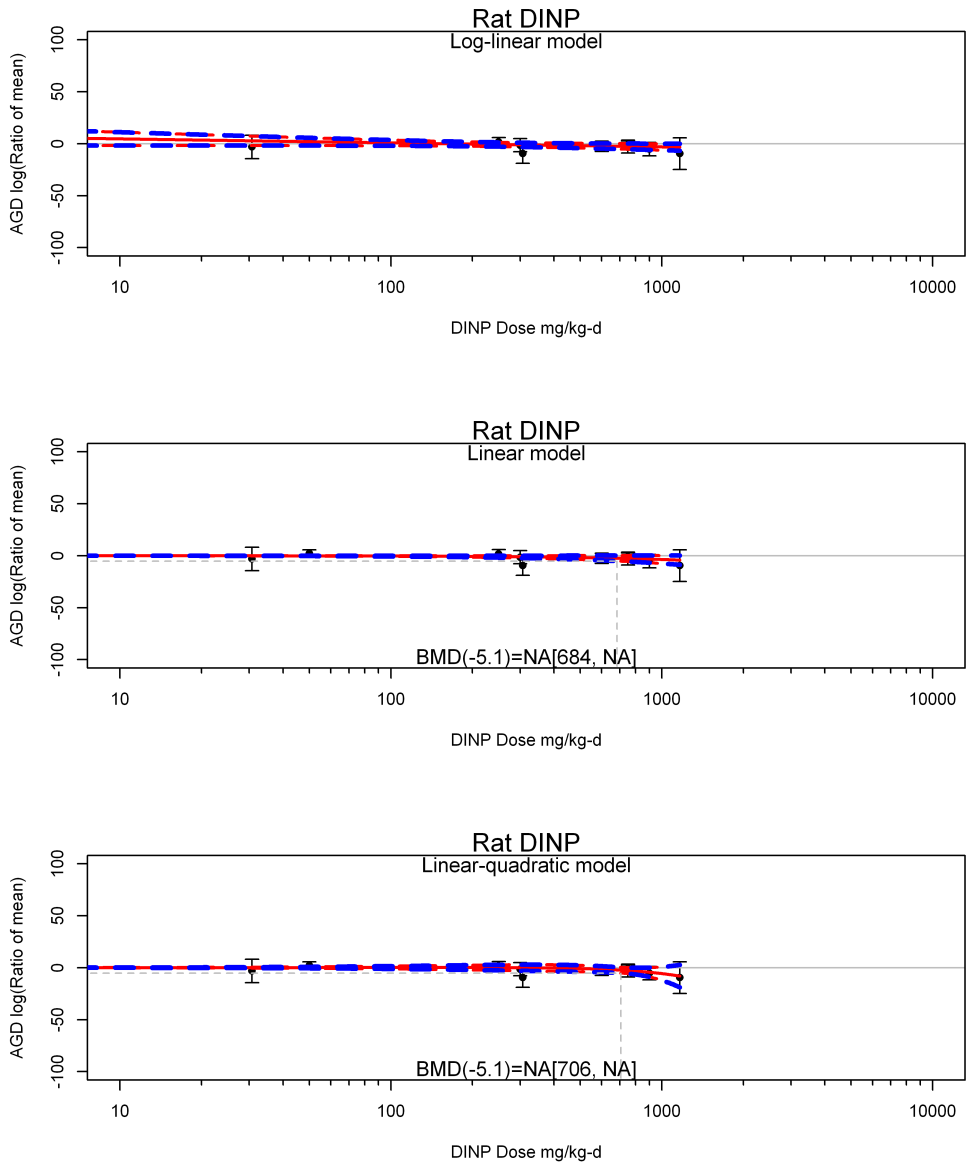


FIGURE C6-14 Benchmark dose estimates from rat studies of DINP and AGD.

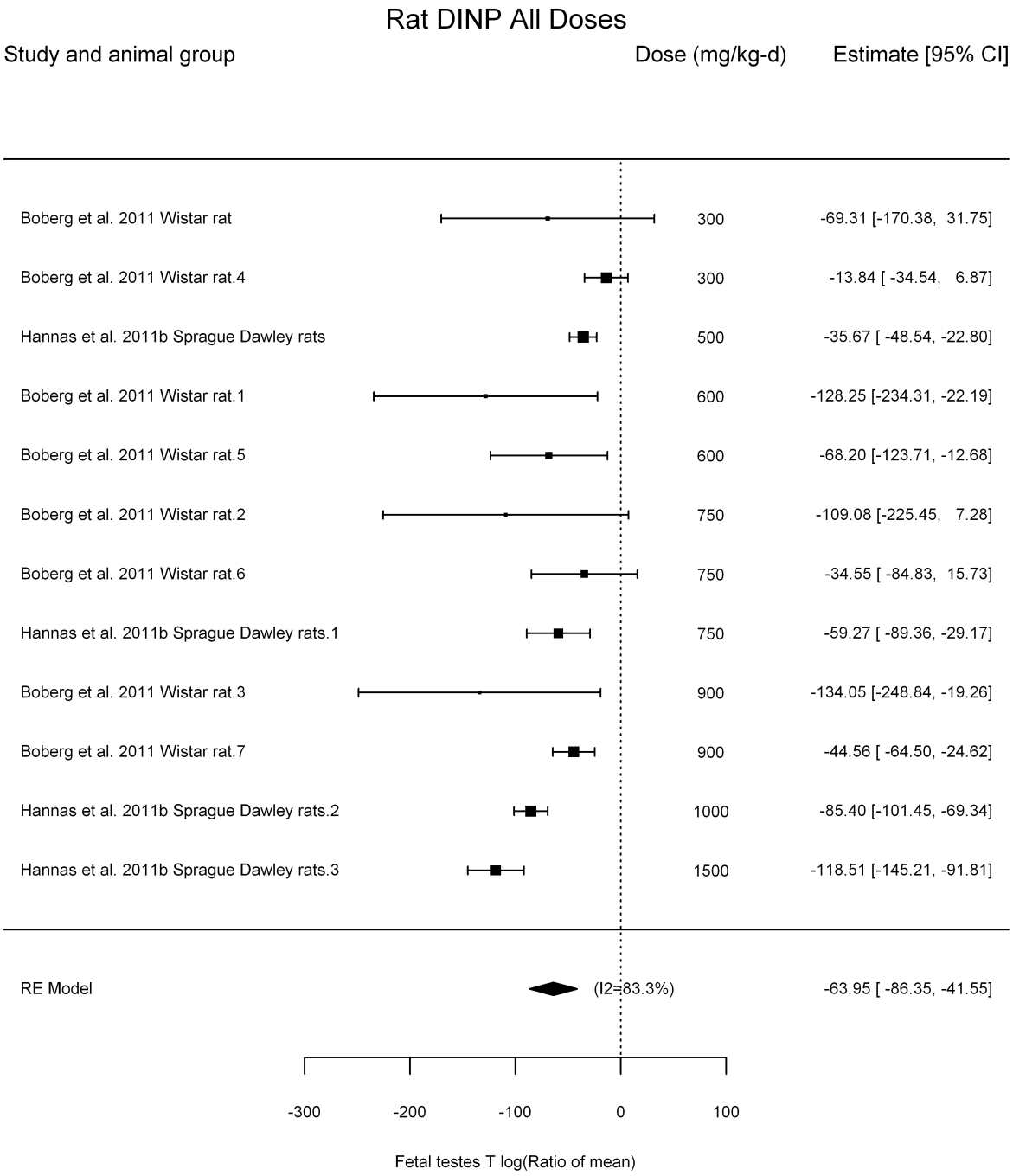


FIGURE C6-15 Meta-analyses of studies of DINP and fetal testosterone in rats.

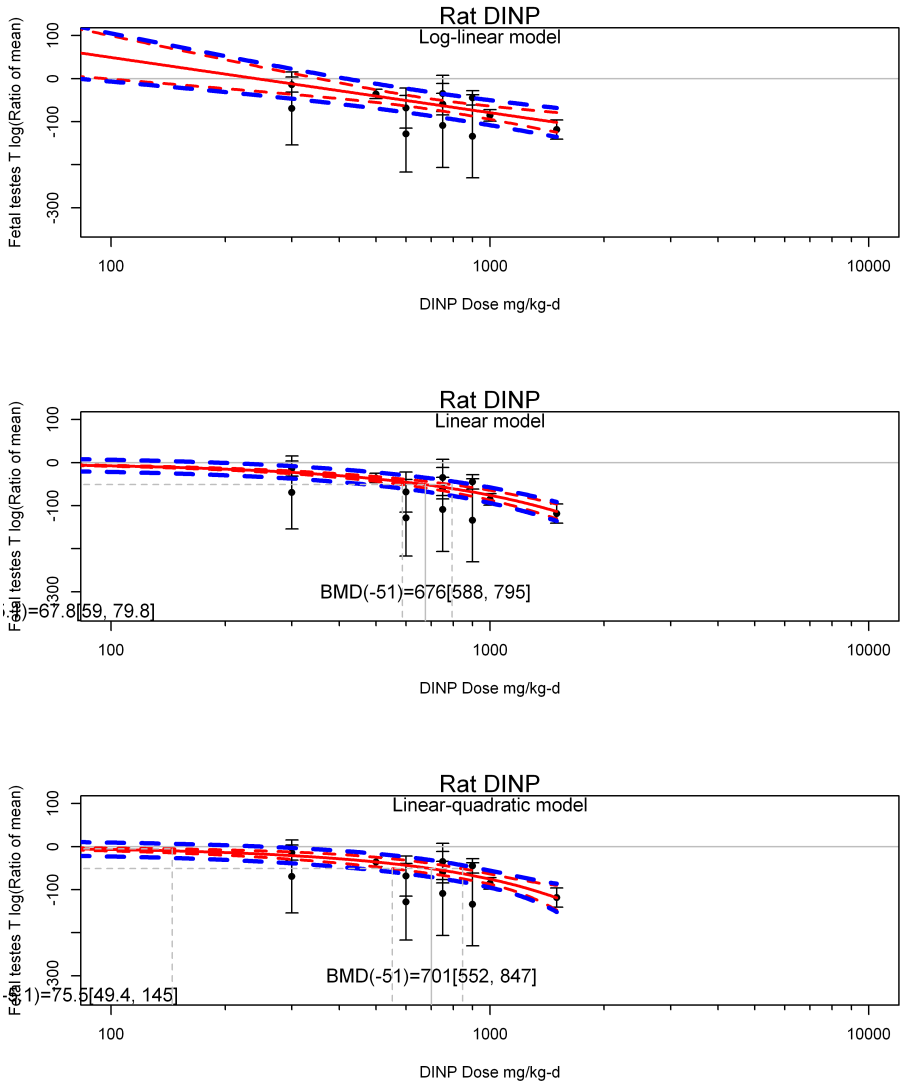


FIGURE C6-16 Benchmark dose estimates from rat studies of DINP and fetal testosterone.

REFERENCES

Adamsson, A., V. Salonen, J. Paranko, and J. Toppari. 2009. Effects of maternal exposure to di-isononylphthalate (DINP) and 1, 1-dichloro-2,2-bis(p-chlorophenyl)ethylene (p,p'-DDE) on steroidogenesis in the fetal rat testis and adrenal gland. Reprod. Toxicol. 28(1):66-74.

Ahmad, R., A.K. Gautam, Y. Verma, S. Sedha, and S. Kumar. 2014. Effects of in utero di-butyl phthalate and butyl benzyl phthalate exposure on offspring development and male reproduction of rat. Environ. Sci. Pollut. Res. Int. 21(4):3156-3165.

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(1):64-74.

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. Reprod. Toxicol. 18(1):53-61.

Borch, J., S.B. Metzdorff, A.M. Vinggaard, L. Brokken, and M. Dalgaard. 2006. Mechanisms underlying the anti-androgenic effects of diethylhexyl phthalate in fetal rat testis. Toxicology 223(1-2):144-155.

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. Environ. Health Perspect. 117(12):1839-1846.

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. Reprod. Toxicol. 30(2):313-321.

Clewell, R.A., J.J. Kremer, C.C. Williams, J.L. Campbell, M.A. Sochaski, M.E. Andersen, and S.J. Borghoff. 2009. Kinetics of selected di-n-butyl phthalate metabolites and fetal testosterone following repeated and single administration in pregnant rats. Toxicology 255(1-2):80-90.

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. Biol. Reprod. 78(6):1018-1028.

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. Reprod. Toxicol. 34(4):614-621.

Ema, M., and E. Miyawaki. 2002. Effects on development of the reproductive system in male offspring of rats given butyl benzyl phthalate during late pregnancy. Reprod. Toxicol. 16(1):71-76.

Ema, M., E. Miyawaki, and K. Kawashima. 2000. Critical period for adverse effects on development of reproductive system in male offspring of rats given di-n-butyl phthalate during late pregnancy. Toxicol. Lett. 111(3):271-278.

Fujii, S., K. Yabe, M. Furukawa, M. Hirata, M. Kiguchi, and T. Ikka. 2005. A two-generation reproductive toxicity study of diethyl phthalate (DEP) in rats. J. Toxicol. Sci. 30(Spec.):97-116.

Giribabu, N., S.B. Sainath, and P.S. Reddy. 2014. Prenatal di-n-butyl phthalate exposure alters reproductive functions at adulthood in male rats. Environ. Toxicol. 29(5):534-544.

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. Toxicol. Sci. 110(2):411-425.

Hannas, B.R., C.S. Lambright, J. Furr, N. Evans, P.M.D. Foster, E.L. Gray, and V.S. Wilson. 2012. Genomic biomarkers of phthalate-induced male reproductive developmental toxicity: A targeted RT-PCR array approach for defining relative potency. Toxicol. Sci. 125(2):544-557.

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. Reprod. Toxicol. 19(4):505-515.

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. Biol. Reprod. 93(4):92.

Klinefelter, G.R., J.W. Laskey, W.M. Winnik, J.D. Suarez, N.L. Roberts, L.F. Strader, B.W. Riffle, and D.N. Veeramachaneni. 2012. Novel molecular targets associated with testicular dysgenesis induced by gestational exposure to diethylhexyl phthalate in the rat: A role for estradiol. Reproduction 144(6):747-761.

Lehmann, K.P., S. Phillips, M. Sar, P.M. Foster, and K.W. Gaido. 2004. Dose-dependent alterations in gene expression and testosterone synthesis in the fetal testes of male rats exposed to di(n-butyl) phthalate. Toxicol. Sci. 81(1):60-68.

Li, L., T. Bu, H. Su, Z. Chen, Y. Liang, G. Zhang, D. Zhu, Y. Shan, R. Xu, Y. Hu, J. Li, G. Hu, Q. Lian, and R.S. Ge. 2015. In utero exposure to diisononyl phthalate caused testicular dysgenesis of rat fetal testis. Toxicol. Lett. 232(2):466-474.

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. Environ. Toxicol. Pharmacol. 35(1):55-60.

Li, N., X. Chen, X. Zhou, W. Zhang, J. Yuan, and J. Feng. 2015. The mechanism underlying dibutyl phthalate induced shortened anogenital distance and hypospadias in rats. J. Pediatr. Surg. 50(12):2078-2083.

Lin, H., R.S. Ge, G.R. Chen, G.X. Hu, L. Dong, Q.Q. Lian, 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. Proc. Natl. Acad. Sci. U.S.A. 105(20):7218-7222.

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. Biol. Reprod. 80(5):882-888.

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. J. Toxicol. Environ. Health A 71(19):1289-1294.

MacLeod, D.J., R.M. Sharpe, M. Welsh, M. Fisken, H.M. Scott, G.R. Hutchison, A.J. Drake, and S. van den Driesche. 2010. Androgen action in the masculinization programming window and development of male reproductive organs. Int. J. Androl. 33(2):279-287.

Mahood, I.K., H.M. Scott, R. Brown, N. Hallmark, M. Walker, and R.M. Sharpe. 2007. In utero exposure to di(n-butyl) phthalate and testicular dysgenesis: Comparison of fetal and adult end points and their dose sensitivity. Environ. Health Perspect. 115(Suppl. 1):55-61.

McKinnell, C., R.T. Mitchell, M. Walker, K. Morris, C.J.H. Kelnar, W.H. Wallace, and R.M. Sharpe. 2009. Effect of fetal or neonatal exposure to monobutyl phthalate (MBP) on testicular development and function in the marmoset. Human Reprod. 24(9):2244-2254.

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. Environ. Health Perspect. 109(3):229-237.

NTP (National Toxicology Program). 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Pocar, P., N. Fiandanese, C. Secchi, A. Berrini, B. Fischer, J.S. Schmidt, K. Schaedlich, and V. Borromeo**.** 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(2):937-948.

Saillenfait, A.M., J.P. Sabate, and F. Gallissot. 2008. Diisobutyl phthalate impairs the androgen-dependent reproductive development of the male rat. Reprod. Toxicol. 26(2):107-115.

Saillenfait, A.M., J.P. Sabate, and F. Gallissot. 2009. Effects of in utero exposure to di-n-hexyl phthalate on the reproductive development of the male rat. Reprod. Toxicol. 28(4):468-476.

Saillenfait, A.M., A.C. Roudot, F. Gallissot, and J.P. Sabate. 2011. Prenatal developmental toxicity studies on di-n-heptyl and di-n-octyl phthalates in Sprague-Dawley rats. Reprod. Toxicol. 32(3):268-276.

Saillenfait, A.M., J.P. Sabate, A. Robert, B. Cossec, A.C. Roudot, F. Denis, and M. Burgart. 2013a. Adverse effects of diisooctyl phthalate on the male rat reproductive development following prenatal exposure. Reprod. Toxicol. 42:192-202.

Saillenfait, A.M., J.P. Sabate, A. Robert, V. Rouiller-Fabre, A.C. Roudot, D. Moison, and F. Denis. 2013b. Dose-dependent alterations in gene expression and testosterone production in fetal rat testis after exposure to di-n-hexyl phthalate. J. Appl. Toxicol. 33(9):1027-1035.

van den Driesche, S., P. Kolovos, S. Platts, A.J. Drake, and R.M. Sharpe. 2012. Inter-relationship between testicular dysgenesis and Leydig cell function in the masculinization programming window in the rat. PLOS One 7(1):e30111.

Vo, T.T., E.M. Jung, V.H. Dang, K. Jung, J. Baek, K.C. Choi, and E.B. Jeung. 2009. Differential effects of flutamide and di-(2-ethylhexyl) phthalate on male reproductive organs in a rat model. J. Reprod. Dev. 55(4):400-411.

Wolfe, G.W., and K.A. Layton. 2005. Diethylhexylphthalate: Multigenerational Reproductive Assessment by Continuous Breeding when Administered to Sprague-Dawley Rats in the Diet. TRC Study No. 7244-200. NTP-RACB 98-004. TherImmune Research Corporation, Gaithersburg, MD.

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. Biol. Res. 46(2):139-146.

Zhang, Y., X. Jiang, and B. Chen. 2004. Reproductive and developmental toxicity in F1 Sprague-Dawley male rats exposed to di-n-butyl phthalate in utero and during lactation and determination of its NOAEL. Reprod. Toxicol. 18(5):669-676.

Appendix D

Supporting Materials for the Phthalate (Human) Systematic Review

SECTION D-1

PHTHALATE (HUMAN) SYSTEMATIC REVIEW PROTOCOL

August 3, 2016
(Modified on October 31, 2016—See Section D-1f)

BACKGROUND AND INTRODUCTION

Phthalates are high production volume chemicals used primarily as plasticizers in many industrial and consumer products. As a result of their ubiquitous use, there is documented widespread human exposure to them. Because the fetus has been shown to be particularly vulnerable to endocrine-disrupting chemicals, such as phthalates, the committee decided to focus on studies of in utero exposure. Ortho-phthalates have been linked to effects on male reproductive-tract development after in utero exposure in human studies.

OBJECTIVE AND SPECIFIC AIMS

Review Question

The overall objective of this systematic review is to answer the question what is the effect of in utero exposure to ortho-phthalates on anogenital distance, hypospadias, or testosterone concentrations in male humans?

The specific aims of the review include

- Identify literature reporting the effects of in utero phthalate exposure on male anogenital distance, hypospadias, or testosterone in humans.
- Extract data on the effects of in utero phthalate exposure on male anogenital distance, hypospadias, or testosterone from relevant studies.
- Assess the internal validity (risk of bias) of individual studies.
- Summarize the extent of evidence available.
- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration such as study-design heterogeneity.
- Rate the confidence in the body of evidence for studies in humans according to one of five statements: (1) high; (2) moderate; (3) low; (4) very low/no evidence available; or (5) evidence of lack of effects on male reproductive-tract development.

PECO Statement

A PECO (Population, Exposure, Comparator, and Outcome) statement was developed by the review team as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the review question for the systematic review.

Population: Male humans

Exposure:

- In utero exposure to any of the following ortho-phthalates or the corresponding monoester or oxidative metabolites: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or dipentyl phthalate (CAS no. 131-18-0).
- No restrictions based on route of exposure. Measurements must be based on biomonitoring data (e.g., urinary monoester or oxidative metabolites, amniotic fluid oxidative phthalate metabolites, oxidative metabolites in other matricies).

Comparator: Male humans exposed in utero to lower concentrations of phthalates.

Outcomes:

- Anogenital distance (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:
  - Anogenital index (AGI): AGD measurement divided by body weight or by the cube root of body weight.
  - Anoscrotal distance (ASD): the measured distance between the anus and base of the scrotum.
  - Anopenile distance (APD): the measured distance from the anus to the base of the penis.
- Hypospadias (incidence, prevalence, and severity/grade) based on clinical guidelines for assessment.
- Testosterone concentrations measured during gestation or at delivery.

METHODS

Problem Formulation and Protocol Development

The review question and specific aims were developed and refined through a series of problem formulation steps. The committee considered review articles on endocrine disruptors in surveying the types of chemicals that might make good case examples, and held a workshop to explore potential case examples, including phthalates. The committee sought an example of a chemical for which both the human and the animal evidence on effects appear to be associated with different exposure levels of that chemical and due to perturbation of the estrogen or androgen hormone system. Phthalates appear to fit this case criterion, and positive feedback was received at the committee's workshop.

Alterations in male reproductive-tract development are the most sensitive effects from exposure to phthalates (NRC 2008). Because the period during in utero sexual differentiation (i.e., the masculinization programming window) is the most sensitive life stage, the exposure period of interest for the systematic

review is in utero. This systematic review will focus on the same end points chosen for the phthalate (animal) systematic review: end points reflecting androgen-dependent adverse effects (AGD and hypospadias), an adverse effect that occurs at relatively low doses (AGD), and a key event in the adverse outcome pathway leading to reduced AGD and hypospadias (fetal testosterone).

Consideration was given to including cryptorchidism as an end point, but the committee decided against it. The mode of action for phthalate-induced cryptorchidism involves reductions in INSL-3 levels in addition to androgen-dependent mechanisms. Important for the committee's charge, there are few, if any, human studies on dose-response relationships between phthalate exposure and cryptorchidism to compare to animal data. Furthermore, studies have shown that rats exposed to phthalates have similar sensitivity to decreased fetal testosterone and AGD as they do for decreased INSL-3, and that cryptorchidism is a less sensitive end point compared to reductions in AGD. Because the overall objective of the committee is to use this systematic review with the one being conducted on the animal evidence to evaluate the coherence between effects and dose-response relationships, the committee judged that it would not be useful to include cryptorchidism in the systematic reviews on phthalates.

The protocol will be peer reviewed by subject-matter and systematic-review experts in accordance with standard report-review practices of the National Academies of Sciences, Engineering, and Medicine. The protocols will be revised in response to peer review comments and will subsequently be published as appendices to the committee's final report. The identity of the peer reviewers will remain anonymous to the committee until the publication of the final report, when their names and affiliations are disclosed in the Preface.

Committee and Staff

There are 11 committee members, supported by two staff members of the National Academies. The committee members were appointed in accordance with the standard policies and practices of the National Academies on the basis of their expertise in general toxicology, reproductive toxicology, developmental toxicology, endocrinology, neurotoxicology, epidemiology, risk assessment, biostatistics, and systematic-review methods. The membership of the committee and the staff was determined before the topic of the systematic review was selected. It was known, however, that each case study would be on an endocrine-disrupting chemical, so committee members who have relevant expertise were specifically recruited and appointed.

Review Team

The review team for this case study will be a subgroup of the committee (RH, SS), two National Academies staff members (EM, SM), and an information specialist (JB). If a member of the review team was a coauthor of a study under review, that member will recuse himself or herself from the evaluation of the quality of that study.

The review team will be responsible for performing all aspects of the review, including conducting the literature searches; applying inclusion/exclusion criteria to screen studies; extracting data; assessing risk of bias for included studies; and analyzing and synthesizing data. The roles and responsibilities of the team members will be documented throughout the protocol. Throughout the course of its work, the review team will also engage other members of the committee to provide consultation needed. The involvement of those individuals will be documented and acknowledged.

Biographical information on the review team is presented in Section D-1a.

Search Methods

Search for Existing Systematic Reviews

The review team will consider using existing systematic reviews to address or help to address its research question. English-language systematic reviews conducted within the last 3 years will be sought. The review team will incorporate prior reviews, update prior reviews, and/or use the reviews as part of its searching, depending on determination of their relevancy and quality (Whitlock et al. 2008). Current guidance on using existing systematic reviews will be used (Robinson et al. 2014, 2015, 2016).

Search

Recent, relevant high-quality systematic reviews addressing the research question about phthalates and male reproductive-tract development will be searched. PubMed will be searched by adding the qualifier "systematic review"[ti] OR "meta-analysis"[pt] OR "meta-analysis"[ti] OR ("systematic"[ti] AND "review"[ti]) OR (systematic review [tiab] AND review [pt]) OR "meta synthesis"[ti] OR "meta synthesis"[ti] OR "integrative review"[tw] OR "integrative research review"[tw] OR "cochrane database syst rev"[ta] OR "evidence synthesis"[tiab] to the preliminary search strategy (see Section D-1b). Language and date restrictions will be applied (English language; published 2013 to present). The systematic review protocol registry PROSPERO (CRD) will also be searched using key terms from the preliminary PubMed strategy.

Study Selection

Two team members (SM, EM) will independently screen search results, applying the following exclusion criteria:

- Not a systematic review.[1] The minimum criteria for a study to be considered a systematic review are
  - conduct of an explicit and adequate literature search,
  - application of predefined eligibility criteria,
  - consideration of the quality of included studies or risk of bias assessment, and
  - synthesis (or attempt at synthesis) of the findings, either qualitatively or quantitatively.
- Not in English.
- Search date prior to 2013.
- Does not match our research question or PECO elements.

For PubMed results, screening will be conducted first using abstracts and then at the full-text level. Results from PROSPERO will be conducted at one level, using the information in the registry. Disagreements regarding eligibility will be resolved through discussion or, where necessary, by a third team member.

Assessment for Quality

Two investigators (KR, AR) will independently assess the risk of bias of eligible systematic reviews using ROBIS (Whiting et al. 2016). Disagreements in rating will be resolved through discussion or, where necessary, through consultation with a third team member. Systematic reviews rated as low quality will be excluded from further consideration at this stage.

[1]A systematic review "is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

Use of Existing Reviews

Eligible systematic reviews of high quality will be reviewed, considering date of search and match with the PECO statement as well as availability of data from the primary studies, how risk of bias was conducted, and other factors. Current reviews considered a good match will be used to address the research question. Reviews that are a good match but with search dates more than a year ago will be updated. If no relevant systematic reviews are found, an independent systematic review will be performed.

Literature Search for Independent Systematic Review

The review team will collaborate with an information specialist (JB) who has training, expertise, and familiarity with developing and performing systematic review literature searches. A variety of methods will be used to identify relevant data (see below). Literature searches will not be limited by publication date.

Online Databases

Electronic searches of the following three online databases will be performed using the search terms outlined in Section D-1b: PubMed, Embase, and Toxline. The search strategy and search terms will be developed by the information specialist (JB), who will implement the search for relevant studies.

Other Resources

Hand searching the reference lists of all the included studies after full-text review will be conducted using the same study selection process as used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

Study Selection

All search results will be imported or manually entered into EndNote (Version x7) reference management software. EndNote will be used to eliminate any duplicate citations before evaluating the eligibility of the citations.

Screening Process

References retrieved from the literature search will be screened for relevance and eligibility against the evidence selection criteria using DistillerSR (Evidence Partners; https://www.evidencepartners.com). Screeners from the review team will be trained with an initial pilot phase on 25 studies undertaken to improve clarity of the evidence selection criteria and to improve accuracy and consistency among screeners. Screening forms are presented in Section D-1c.

Title and Abstract Screening

Each citation will be independently screened by two reviewers (SM, EM) to determine whether it meets the selection criteria for inclusion that reflect the PECO statement with some additional considerations as listed below. Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

The title/abstract screening form will be used to screen and EXCLUDE references if at least one of the following criteria is met:

1. No original data (e.g., review article, commentary, editorial)

2. Study does not include male humans
3. Study does not report phthalate exposure
4. No relevant outcomes
5. Incomplete information (e.g., conference abstract, meeting poster)
6. Not in English and unable to determine eligibility
7. Other (explanation required)

The following types of records will be INCLUDED at the title/abstract level: any English-language study of male humans exposed to phthalates in utero.

Only English-language publications will be included, because of time and resource constraints. There appears to be no indication that foreign-language publications would make a contribution that is distinct from what is found in the English-language literature.

Updated details to instructions and interpretations for title and abstract screening will be added to the Section D-1f to document the process of the review team during the screening process.

Full-Text Screening

Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers involved in title and abstract screening (SM, EM). Each reference will be screened in duplicate and independently. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

Citations will be EXCLUDED at the full-text level if at least one of the following criteria is met:

1. No original data (e.g., review article, commentary, editorial)
2. Study does not include male humans
3. Study does not report phthalate exposure to one or more of the phthalates listed in the PECO statement
4. Study does not have biomonitoring data specific to phthalate exposure
5. Study does not include in utero exposure
6. Study does not assess or report anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or testosterone concentrations measured during gestation or at delivery
7. No comparator group (males exposed in utero at lower concentrations of phthalates)
8. Not in English
9. Other reason (explanation required)

The reason for exclusion at the full-text-review stage will be annotated and reported in a study selection flow diagram in the final report (following PRISMA [Moher et al. 2009]). The reasons for exclusion will be documented from the list (1-9) above.

Citations will be INCLUDED if they meet the PECO statement criteria:

- Study includes male humans
- Study includes in utero exposure
- Study includes comparison with males exposed in utero at lower concentrations
- Study measures anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or testosterone concentrations

Updated details to instructions and interpretations for full-text screening will be added to the Section D-1f to document the process of the review team during the screening process.

Data Extraction

Data will be collected and recorded (extracted) from included studies by one member of the review team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction will be resolved through discussion. The extracted data will be used to summarize study designs and findings and/or to conduct statistical analyses. Section D-1d presents the data extraction elements that will be used.

The review team will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The study extraction files will note whether an attempt was made to contact study authors by email for missing data considered important for evaluating key study findings (and whether or not a response was received).

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 review will include all publications on the study, select one publication to use as the primary publication, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up or, for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. The review will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, the review team will include a single instance of the data (and avoid more than one—that is, duplicate instances of the data).

Data extraction will be completed using the Health Assessment Workspace Collaborative (HAWC) software, an open source and freely available Web-based interface application, for visualization and warehousing.[2]

Risk of Bias (Quality) Assessment of Individual Studies

Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect. Internal validity or risk of bias will be assessed for individual studies using a tool developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) that outlines an approach to evaluating risk of bias for human epidemiology studies. The risk of bias domains and questions are based on established guidance for observational human studies and randomized controlled trials (Higgins and Green 2011; Viswanathan et al. 2012, 2013; Sterne et al. 2014). The risk of bias tool includes a common set of questions (Section D-1e) that are answered based on the specific details of individual studies to develop risk of bias ratings (using the four options: definitely low risk of bias; probably low risk of bias; probably high risk of bias; or definitely high risk of bias). Study design determines the subset of questions that should be used to assess risk of bias for an individual study (see Table D1-1).

Studies are independently assessed by two assessors (RH, SS) who answer all applicable risk of bias questions with one of four options (see Table D1-2) following prespecified criteria detailed in Section D-1e. The criteria describe aspects of study design, conduct, and reporting required to reach risk of bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific type of human study designs. Risk of bias will be assessed at the outcome level because study

[2]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

design or method specifics may increase the risk of bias for some outcomes and not others within the same study.

Information or study procedures that were not reported are assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Authors will be queried by email to obtain missing information, and responses received were used to update risk of bias ratings.

Assessors will be trained in using the criteria to develop risk of bias ratings for each question, with an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk of bias assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk of bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk of bias criteria.

After assessors have independently made risk of bias determinations for a study across all risk of bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered and resolved with the review team. The final risk of bias rating for each question will be recorded along with a statement of the basis for that rating.

Data Analysis and Evidence Synthesis

The review team will qualitatively synthesize the body of evidence for each outcome and, where appropriate, a meta-analysis will be performed. If a meta-analysis is performed, summaries of main characteristics for each included study will be compiled and reviewed by two team members to determine comparability between studies, to identify data transformations necessary to ensure comparability, and to determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups (e.g., quartiles of exposure)
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring)
- Measurement of biomonitoring data specific to phthalate exposure for each exposure group
- Health outcome(s) reported
- Conditioning variables in the analysis (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

The review team expects to require input from subject-matter experts to help assess the heterogeneity of the studies. Subgroup analyses to examine the extent to which risk of bias contributes to heterogeneity will be performed. Situations where it may not be appropriate to include a study are when data on exposure or outcome are too different to be combined or other circumstances that may indicate that averaging study results would not produce meaningful results. When considering outcome measures for conducting meta-analyses, continuous outcome measures, such as beta coefficients (and their associated confidence intervals) from regression analysis, are preferred. A secondary alternative, when there are more than two groups, is to conduct a regression analysis of the odds or risk ratios across exposure groups and to use the derived beta coefficient. A tertiary alternative when there are only two groups (e.g., higher and lower exposure) is to use the odds or risk ratio itself.

TABLE D1-1 OHAT Risk of Bias Tool

Risk-of-Bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Nonhuman animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.
**Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, nonrandomized experimental studies).
***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).
SOURCE: NTP (2015, p. 37).

TABLE D1-2 Answers to the Risk of Bias Questions

++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices.
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias.
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk–of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices.

SOURCE: NTP (2015, p. 36).

If a meta-analysis is conducted, a random effects model will be used for the analysis. Heterogeneity will be assessed using the I-squared statistic. Interpretation of I-squared will be based on the Cochrane Handbook: 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). Additionally, as described in the Cochrane Handbook, for the last three categories, the importance of the I-squared will be interpreted considering not only the magnitude of effects but also the strength of the evidence (90% two-tailed confidence interval).

The review team will also perform sensitivity analyses on the following aspects:

- Sensitivity to exclusion of individual studies in succession,
- Sensitivity to alternative exposure metrics (if available), and
- Sensitivity to alternative outcome metrics (if available).

It is unlikely that there will be enough studies or information to meaningfully assess publication bias or to perform subgroup analyses, so no such analyses are planned.

In the event that these proposed methods for data analysis are altered to tailor to the evidence base from included studies, the protocol will be amended accordingly, and the reasons for change will be justified in the documentation.

Confidence Rating: Assessment of the Body of Evidence

The quality of evidence for each male reproductive outcome will be evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in NTP (2015, see Step 5). In brief, available studies on a particular outcome are initially grouped by key study-design features, and each grouping of studies is given an initial confidence rating by those features.

The initial rating is downgraded for factors that decrease confidence in the results, including

- high risk of bias
- unexplained inconsistency
- indirectness or lack of applicability
- imprecision
- publication bias

The initial rating is upgraded for factors that increase confidence in the results, including

- large magnitude of effect
- dose-response relationship
- consistency across study designs/populations/animal models or species
- consideration of residual confounding
- other factors that increase our confidence in the association or effect (e.g., particularly rare outcomes)

The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt et al. 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney et al. 2014).

Confidence ratings are independently assessed by members of the review team, and discrepancies will be resolved by consensus and consultation with technical advisors as needed. Confidence ratings will be summarized in evidence profile tables.

REFERENCES

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

Higgins, J., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available http://handbook.cocharne.org [accessed May 6, 2016].

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. J. Clin. Epidemiol. 62(10):1006-1012.

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-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Robinson, K.A., E.P. Whitlock, M.E. O'Neil, J.K. Anderson, L. Hartling, D.M. Dryden, M. Butler, S.J. Newberry, M. McPheeters, N.D. Berkman, J.S. Lin, and S. Chang S. 2014. Integration of Existing Systematic Reviews. Research white paper. AHRQ Publication No. 14-EHC016-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK216379/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2015. Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 15-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279904/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2016. Twelve recommendations for integrating existing systematic reviews into new reviews: EPC guidance. J. Clin. Epidemiol. 70:38-44.

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. Environ. Health Perspect. 122(7):711-718.

Sterne, J.A.C., J.P.T. Higgins, and B.C. Reeves. 2014. ACROBAT-NRSI: A Cochrane risk of Bias Assessment Tool for Non-randomized Studies of Interventions. Version 1.0.0. September 24, 2014 [online]. Available: www.riskofbias.info [accessed May 6, 2016].

Viswanathan, M., M. Ansari, N.D. Berkman, S. Chang, L. Hartling, L.M. McPheeters, P.L. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze, and J.R. Treadwell. 2012. Assessing the Risk of Bias of Individual Studies when Comparing Medical Interventions. AHRQ Publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available: www.effectivehealthcare.ahrq.gov/ [accessed May 6, 2016].

Viswanathan, M., N.D. Berkman, D.M. Dryden, and L. Hartling. 2013. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank. Methods Research Report. AHRQ Publication No. 13-EHC106-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available: www.effectivehealthcare.ahrq.gov/reports/final.cfm [accessed May 6, 2016].

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

Whitlock, E.P., J.S. Lin, R. Chou, P. Shekelle, and K.A. Robinson. 2008. Using existing systematic reviews in complex systematic reviews. Ann. Intern. Med. 148(10):776-782.

SECTION D-1a

REVIEW TEAM BIOGRAPHICAL INFORMATION

Jaime F. Blanck is a clinical informationist at the Welch Medical Library at Johns Hopkins University. She creates and implements systematic review search strategies across multiple databases and provides comprehensive reference, research, and information services to multiple departments within the School of Medicine. She received an MLIS from the University of Pittsburgh and an MPA from the University of Baltimore.

Russ B. Hauser is the Frederick Lee Hisaw Professor of Reproductive Physiology and Professor of environmental and occupational epidemiology in the Department of Environmental Health at the Harvard T.H. Chan School of Public Health. He also holds an appointment at the Harvard Medical School, where he is professor of obstetrics, gynecology, and reproductive biology. Dr. Hauser's research focuses on the health risks posed by exposure to environmental chemicals that adversely affect human development and reproductive health. He has served on several NRC and IOM committees, including the Committee to Review EPA's State of the Science Paper on Nonmonotonic Dose Response and the Committee on the Health Risks of Phthalates. Dr. Hauser is a member of two EPA Science Advisory Boards. He served on the US Consumer Product Safety Commission's Chronic Hazard Advisory Panel examining the effects of phthalates on children's health. He received an MD from the Albert Einstein College of Medicine and an MPH and a ScD from the Harvard School of Public Health.

Ellen Mantus is a scholar and director of risk assessment on the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine with more than 20 years of experience in the fields of toxicology and risk assessment. She has served as the study director on numerous projects, including ones that have assessed the health implications of various chemical exposures; developed strategies for applying modern scientific approaches in toxicology and risk assessment; provided guidance to federal agencies on risk-based decision making; and evaluated barriers to deployment of electric vehicles and associated charging infrastructure. Before joining the National Academies, Dr. Mantus was a project manager with ICF Consulting where she served as a primary reviewer for numerous toxicological studies and provided risk assessment and regulatory support on a wide array of projects. Dr. Mantus received a PhD in chemistry from Cornell University.

Susan Martel is a senior program officer in the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine. She has 20 years of experience in supporting toxicology and risk assessment projects for the US Environmental Protection Agency, the US Department of Defense, and the National Aeronautics and Space Administration. Recent projects include working with committees evaluating the toxicological effect of arsenic, developing exposure guidelines for use on spacecraft, and assessing pesticide risks-assessment practices. Before joining the National Academies, she was the administrator of the Registry for Toxicology Pathology for Animals at the American Registry of Pathology. She received a BA in biology from Skidmore College.

Andrew A. Rooney is deputy director of the Office of Health Assessment and Translation (OHAT) in the National Toxicology Program at the National Institute of Environmental Health Sciences. He has been developing risk assessment methods and guidance throughout his professional career and is a principal author of the 2012 WHO/IPCS Guidance for Immunotoxicity Risk Assessment for Chemicals. Most recently, he has been working on emerging issues in toxicology and environmental health, including methods to address study quality in terms of risk of bias for human, animal, and mechanistic studies and adaptation of systematic review methods for addressing environmental health questions. He led the team that developed the OHAT approach to systematic review. Dr. Rooney has an MS and a PhD in zoology from the University of Florida.

Sheela Sathyanarayana is an associate professor in the Department of Pediatrics and an Adjunct Associate Professor in the Department of Environmental and Occupational Health Sciences at the University of Washington. She is also an attending physician at Harborview Medical Center and Seattle Children's Hospital. Her research interests focus on exposures to endocrine disrupting chemicals, including phthalates and bisphenol A, and their effects on reproductive development. Currently, Dr. Sathyanarayana is the center director and clinical director for The Infant Development and Environment Study, which is a multicenter cohort study of phthalate exposures in pregnancy and health outcomes in children. She also chairs EPA's Children's Health Protection Advisory Committee. Dr. Sathyanarayana earned an MD from the University of Southern California and an MPH in epidemiology from the University of Washington.

SECTION D-1b

LITERATURE SEARCH STRATEGY

The review team will employ a multi-method process to identify all potentially relevant studies as detailed below.

Electronic Searches

PubMed

A search string employing medical subject heading (MeSH) terms and keyword synonyms will be developed. The PubMed search strategy will be considered the primary search strategy and will provide the basis of the other electronic search strategies. To assist in compiling these terms, the review team will conduct a text analysis of eight articles known to the authors. These articles were selected because they represent both American and non-American publications and will help identify spelling variants. The search strategies will address each of the following concepts:

- *Phthalates*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the following phthalates: the CAS numbers to these 11 phthalates: benzylbutyl phthalate (CAS no. 85-68-7), dibutyl phthalate (CAS no. 84-74-2), diethyl phthalate (CAS 84-66-2), diethylhexyl phthalate (CAS no. 117-81-7), diisobutyl phthalate (CAS no. 84-69-5), diisononyl phthalate (CAS no. 28553-12-0), diisooctyl phthalate (CAS no. 27554-26-3), dimethyl phthalate (CAS no. 131-11-3), di-n-octyl phthalate (CAS no. 117-84-0), diisodecyl phthalate (CAS no. 26761-40-0), and/or di-pentyl phthalate (CAS no. 131-18-0). The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. CAS registry numbers for each phthalate substance will also be included in the list of search terms. All MeSH terms, Supplementary Concept terms, keyword synonyms, and CAS registry numbers will be searched together as one concept using the Boolean operator "OR."
- *Exposure*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the *exposure* concept. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. All MeSH terms and keyword synonyms will be searched together as one concept using the Boolean operator "OR."
- *Human studies*—The search filter developed by the Cochrane Library to identify human studies (see http://handbook.cochrane.org/ part 2, section 6.4.f) will be modified to comply with PubMed formatting.
- *Outcomes*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to male genital abnormalities.

Each of the above concepts will be searched together using the Boolean operator "AND." There will not be limitations on date of publication, language, or publication type. All citation records will be exported to EndNote. Additional citations identified through the search processes identified below will also be exported to the project EndNote library. Duplicates will be removed from the citation library using the "Find Duplicates" tool in EndNote as well as a manual review of citations by the project librarian to identify any duplicates not found during the automated process. The number of citations found in each database will be recorded, as well as the number of duplicates and final tally of unique citations. The final

library of citations will be uploaded to the Health Assessment Workspace Collaboration Web-based tool (www.hawcproject.org) for systematic reviews where they will be reviewed by the team.

Embase

The controlled vocabulary database Emtree is used by Embase. For each MeSH term identified through the process above, Emtree will be searched for the appropriate corresponding term. Additional keywords will be identified using the list of synonyms from each Emtree record and added to the keywords from the MeSH records.

Toxline

The review team will develop the Toxline search strategy by removing any database specific formatting from the PubMed search strategy to create a keyword-only search (Toxline does not employ a controlled vocabulary).

Search Strategies

PubMed

("butylbenzyl phthalate" [Supplementary Concept] OR "Dibutyl Phthalate"[Mesh] OR "diethyl phthalate" [Supplementary Concept] OR "Diethylhexyl Phthalate"[Mesh] OR "diisobutyl phthalate" [Supplementary Concept] OR "diisononyl phthalate" [Supplementary Concept] OR "diisooctyl phthalate" [Supplementary Concept] OR "dimethyl phthalate" [Supplementary Concept] OR "di-n-octyl phthalate" [Supplementary Concept] OR "benzylbutyl phthalate"[tw] OR "benzyl butyl phthalate"[tw] OR "butyl benzyl phthalate"[tw] OR "butylbenzyl phthalate"[tw] OR "butylbenzylphthalate"[tw] OR "phthalic acid butyl benzyl ester"[tw] OR "butyl-benzyl-phthalate"[tw] OR "BBzP"[tw] OR "BzBP"[tw] OR "BBPHT"[tw] OR "85-68-7"[tw] OR "Dibutyl Phthalate"[tw] OR "Di-n-Butyl Phthalate"[tw] OR "Di n Butyl Phthalate"[tw] OR "Butyl Phthalate"[tw] OR "d n butyl phthalate"[tw] OR "dbp"[tw] OR "di n butyl phthalate"[tw] OR "dibutyl phthalate"[tw] OR "dibutylphthalate"[tw] OR "phthalic acid di n butyl este"[tw] OR "84-74-2"[tw] OR "phthalic acid diethyl ester"[tw] OR "diethyl phthalate"[tw] OR "diethylphthalate"[tw] OR "ethyl phthalate"[tw] OR "di-ethyl phthalate"[tw] OR "DEP"[tw] OR "84-66-2"[tw] OR "bis (2 ethylhexyl) phthalate"[tw] OR "bis (2 ethylhexyl) phthalate"[tw] OR "bis (2 ethylhexyl) phthalate"[tw] OR "bis (2 ethylhexylphthalate)"[tw] OR "Bis(2-ethylhexyl)phthalate"[tw] OR "DEHP"[tw] OR "di (2 ethylhexyl) phthalate"[tw] OR "di 2 ethylhexyl phthalate"[tw] OR "di 2 ethylhexylphthalate"[tw] OR "Di-2-Ethylhexylphthalate"[tw] OR "diethylhexyl phthalate"[tw] OR "Di-octyl Phthalate"[tw] OR "octoil"[tw] OR "phthalic acid di 2 ethylhexyl ester"[tw] OR "phthalic acid diethylhexyl ester"[tw] OR "117-81-7"[tw] OR "di-iso-butyl phthalate"[tw] OR "DiBP"[tw] OR "84-69-5"[tw] OR "di-isononylphthalate"[tw] OR "ENJ 2065"[tw] OR "ENJ-2065"[tw] OR "di-isononyl phthalate"[tw] OR "di-iso-nonyl phthalate"[tw] OR "DINP"[tw] OR "28553-12-0"[tw] OR "Diisooctylphthalate"[tw] OR "27554-26-3"[tw] OR "diamyl phthalate"[tw] OR "dipentyl phthalate"[tw] OR "phthalic acid dipentyl ester"[tw] OR "dipentyl benzene-1,2-dicarboxylate"[tw] OR "di-n-pentyl phthalate"[tw] OR "131-18-0"[tw] OR "Dimethyl phthalate"[tw] OR "Dimethylphthalate"[tw] OR "Avolin"[tw] OR "Citrola"[tw] OR "Dmp"[tw] OR "dmp30"[tw] OR "fermine"[tw] OR "methyl phthalate"[tw] OR "mipax"[tw] OR "mugia"[tw] OR "palatinol m"[tw] OR "sketofax"[tw] OR "131-11-3"[tw] OR "Di-n-octyl phthalate"[tw] OR "di n octyl phthalate"[tw] OR "di n octylphthalate"[tw] OR "dioctyl phthalate"[tw] OR "dioctylphthalate"[tw] OR "di(n-octyl)phthalate"[tw] OR "phthalic acid di n octyl ester"[tw] OR "DNOP"[tw] OR "117-84-0"[tw]) AND ("Maternal Exposure"[Mesh] OR "Environmental Exposure"[Mesh:NoExp] OR "Prenatal Exposure Delayed Effects"[Mesh] OR "Exposure"[tw] OR "Exposed"[tw] OR "exposures"[tw] OR "exposing"[tw] AND ("Genital Diseases, Male"[Mesh] OR "Genitalia, Male"[Mesh] OR "Testosterone"[Mesh:NoExp] OR "Androgens"[Mesh] OR "Anogeni-

tal"[tw] OR "AGD"[tw] OR "AGI"[tw] OR "ASD"[tw] OR "APD"[tw] OR "Urogenital"[tw] OR "Penile"[tw] OR "penis"[tw] OR "Anoscrotal"[tw] OR "Anopenile"[tw] OR "anorectal"[tw] OR "Testosterone"[tw] OR "androgen"[tw] OR "androgens"[tw] OR "Hypospadias"[tw] OR "hypospadia"[tw] OR "Testis"[tw] OR "testes"[tw] OR (("Anorectal"[tw] OR "genital"[tw] OR "genitals"[tw] OR "testes"[tw] OR "rectum"[tw]) AND ("malformation"[tw] OR "malformations"[tw] OR "development"[tw] OR "abnormalities"[tw] OR "abnormality"[tw] OR "dysplasia"[tw])) OR ("Male"[tw] and ("genital"[tw] OR "genitals"[tw] OR "genitalia"[tw])) OR ("Anus"[tw] AND ("genital"[tw] OR "genitals"[tw] OR "genitalia"[tw]))) NOT (("Animals"[Mesh] NOT ("Animals"[Mesh] AND ("Humans"[Mesh]))))

Embase

"phthalic acid benzyl butyl ester"/exp OR "phthalic acid dibutyl ester"/exp OR "phthalic acid diethyl ester"/exp OR "phthalic acid bis(2 ethylhexyl) ester"/exp OR "phthalic acid dimethyl ester"/exp OR "phthalic acid dioctyl ester"/exp OR "benzylbutyl phthalate" OR "benzyl butyl phthalate" OR "butyl benzyl phthalate" OR "butylbenzyl phthalate" OR "butylbenzylphthalate" OR "phthalic acid butyl benzyl ester" OR "butyl-benzyl-phthalate" OR "BBzP" OR "BzBP" OR "BBPHT" OR "85-68-7" OR "Dibutyl Phthalate" OR "Di-n-Butyl Phthalate" OR "Di n Butyl Phthalate" OR "Butyl Phthalate" OR "d n butyl phthalate" OR "dbp" OR "di n butyl phthalate" OR "dibutyl phthalate" OR "dibutylphthalate" OR "phthalic acid di n butyl este" OR "84-74-2" OR "phthalic acid diethyl ester" OR "diethyl phthalate" OR "diethylphthalate" OR "ethyl phthalate" OR "di-ethyl phthalate" OR "DEP" OR "84-66-2" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexyl) phthalate" OR "bis (2 ethylhexylphthalate)" OR "Bis(2-ethylhexyl)phthalate" OR "DEHP" OR "di (2 ethylhexyl) phthalate" OR "di 2 ethylhexyl phthalate" OR "di 2 ethylhexylphthalate" OR "Di-2-Ethylhexylphthalate" OR "diethylhexyl phthalate" OR "Dioctyl Phthalate" OR "octoil" OR "phthalic acid di 2 ethylhexyl ester" OR "phthalic acid diethylhexyl ester" OR "117-81-7" OR "di-iso-butyl phthalate" OR "DiBP" OR "84-69-5" OR "di-isononylphthalate" OR "ENJ 2065" OR "ENJ-2065" OR "di-isononyl phthalate" OR "di-isononyl phthalate" OR "DINP" OR "28553-12-0" OR "Diisooctylphthalate" OR "27554-26-3" OR "diamyl phthalate" OR "dipentyl phthalate" OR "phthalic acid dipentyl ester" OR "dipentyl benzene-1,2-dicarboxylate" OR "di-n-pentyl phthalate" OR "131-18-0" OR "Dimethyl phthalate" OR "Dimethylphthalate" OR "Avolin" OR "Citrola" OR "Dmp" OR "dmp30" OR "fermine" OR "methyl phthalate" OR "mipax" OR "mugia" OR "palatinol m" OR "sketofax" OR "131-11-3" OR "Di-n-octyl phthalate" OR "di n octyl phthalate" OR "di n octylphthalate" OR "dioctyl phthalate" OR "dioctylphthalate" OR "di(n-octyl)phthalate" OR "phthalic acid di n octyl ester" OR "DNOP" OR "117-84-0" AND ('male genital system disease'/exp OR 'male genital system'/exp OR 'testosterone'/exp OR 'androgen'/de OR "Anogenital":ti,ab OR "AGD":ti,ab OR "AGI":ti,ab OR "ASD":ti,ab OR "APD":ti,ab OR "Urogenital":ti,ab OR "Penile":ti,ab OR "penis":ti,ab OR "Anoscrotal":ti,ab OR "Anopenile":ti,ab OR "anorectal":ti,ab OR "Testosterone":ti,ab OR "androgen":ti,ab OR "androgens":ti,ab OR "Hypospadias":ti,ab OR "hypospadia":ti,ab OR "Testis":ti,ab OR "testes":ti,ab OR (("Anorectal":ti,ab OR "genital":ti,ab OR "genitals":ti,ab OR "testes":ti,ab OR "rectum":ti,ab) AND ("malformation":ti,ab OR "malformations":ti,ab OR "development":ti,ab OR "abnormalities":ti,ab OR "abnormality":ti,ab OR "dysplasia":ti,ab)) OR ("Male":ti,ab and ("genital":ti,ab OR "genitals":ti,ab OR "genitalia":ti,ab)) OR ("Anus":ti,ab AND ("genital":ti,ab OR "genitals":ti,ab OR "genitalia":ti,ab))) AND ('prenatal exposure'/exp OR 'environmental exposure'/exp OR 'exposure' OR 'exposed' OR 'exposures' OR 'exposing' NOT ('animal'/exp NOT ('animal'/exp AND 'human'/exp))

Toxline

(("117-81-7" OR "117-84-0" OR "131-11-3" OR "131-18-0" OR "27554-26-3" OR "28553-12-0" OR "84-66-2" OR "84-69-5" OR "84-74-2" OR "85-68-7" OR "Avolin" OR "BBPHT" OR "BBzP" OR "bis 2 ethylhexylphthalate" OR "Bis 2-ethylhexyl phthalate" OR "butylbenzylphthalate" OR "butyl-benzyl-phthalate" OR "BzBP" OR "Dbp" OR "DEP" OR "di n octylphthalate" OR "DiBP" OR "dieth-

ylphthalate" OR "di-isononylphthalate" OR "Diisooctylphthalate" OR "Dimethylphthalate" OR "DINP" OR "dioctylphthalate" OR "dipentyl benzene-1,2-dicarboxylate" OR "Dmp" OR "dmp30" OR "DNOP" OR "ENJ 2065" OR "fermine" OR "mipax" OR "mugia" OR "octoil" OR "o-phthalate" OR "o-phthalates" OR "palatinol" OR "sketofax") AND ("Exposure" OR "Exposed" OR "exposures" OR "exposing") AND ("Anogenital" OR "AGD" OR "AGI" OR "ASD" OR "APD" OR "Urogenital" OR "Penile" OR "penis" OR "Anoscrotal" OR "Anopenile" OR "anorectal" OR "Testosterone" OR "androgen" OR "androgens" OR "Hypospadias" OR "hypospadia" OR "Testis" OR "testes" OR (("Anorectal" OR "genital" OR "genitals" OR "testes" OR "rectum") AND ("malformation" OR "malformations" OR "development" OR "abnormalities" OR "abnormality" OR "dysplasia")) OR ("Male" and ("genital" OR "genitals" OR "genitalia")) OR ("Anus" AND ("genital" OR "genitals" OR "genitalia"))) NOT (animals OR animal OR mice OR mouse OR rats OR rat OR rodent OR rodents OR fish)

SECTION D-1c

SCREENING FORMS

Title and Abstract Screening Form

Instructions: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include male humans	
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)	

Full-Text Screening Form

Instructions: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include male humans	
Study does not report exposure to one or more of the phthalates listed in the PECO statement	
Study does not have biomonitoring data specific to phthalate exposure	
Study does not include in utero exposure	
Study does not assess or report anogenital distance, anogenital index, anoscrotal distance, anopenile distance, hypospadias, or testosterone concentration measured during gestation or at delivery	
No comparator group (male humans exposed in utero to lower concentrations of phthalates)	
Not in English	
Other (explanation required)	

SECTION D-1d

DATA EXTRACTION ELEMENTS FOR HUMAN STUDIES

Funding	Funding source(s)
	Reporting of conflict of interest (COI) by authors (*reporting bias)
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or life stage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, n per group in analysis, and participation/follow-up rates) (*missing data bias)
	Inclusion/exclusion criteria/recruitment strategy (*selection bias)
	Description of reference group (*selection bias)
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report, etc.)
	Length of follow-up (*information bias)
	Health outcome category (e.g., cardiovascular)
	Health outcome (e.g., blood pressure) (*reporting bias)
	Diagnostic or methods used to measure health outcome (*information bias)
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed) (*confounding bias)
	Substance name or CAS number
	Exposure assessment (e.g., blood, urine, hair, air, drinking water, job classification, residence, administered treatment in controlled study, etc.) (*information bias)
	Methodological details for exposure assessment (e.g., HPLC-MS/MS, limit of detection) (*information bias)
	Statistical methods (*information bias)
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results. When possible, measures of effect will be converted to a common metric with associated 95% confidence intervals (CIs). Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percent control response. Categorical data are typically expressed as odds ratio, relative risk (RR, also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies.
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias.

SECTION D-1e

RISK OF BIAS QUESTIONS FOR EPIDEMIOLOGIC STUDIES

Cohort Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and nonexposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates. • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed, but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and nonexposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and nonexposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group, including a different rate of nonresponse without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and nonexposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

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

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias, including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that 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, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as key/primary potential confounders and/or effect measure modifiers that must be considered in the analyses of the relationship between phthalate exposure and male reproductive outcomes: 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.

- **Note:** The following variables should be considered as additional potential confounders and/or effect measure modifiers, but consideration of these variables is not required in the analysis of the relationship between phthalate exposure and male reproductive outcomes: maternal age, pre-pregnancy or maternal BMI, maternal education, maternal income, maternal race/ethnicity, and time of day of urine collection.

Probably Low Risk of Bias (+)

- Indirect evidence that appropriate adjustments were made,
- **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,
- **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements,
- **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research),
- **OR** it is deemed that co-exposures present would not appreciably bias results.
- **Note:** This includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the distribution of important 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 (record "NR" as basis for answer),
- **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity,
- **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer),
- **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 (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses,
- **OR** there is direct evidence that covariates and confounders considered were assessed using nonvalid measurements,
- **OR** 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.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)

- 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.
- **Note:** 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 (+)

- 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 (-) or (NR)
• 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 (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. • **Note:** 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.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time frame) using well-established methods that directly measure exposure (e.g., measurement of urinary phthalate metabolites [and a measure of urinary dilution was available], amniotic fluid oxidative phthalate metabolites). • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the male reproductive outcome was assessed using well-established methods (e.g., gold standard), • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes.

• **Note:** Well-established methods will depend on the outcome, and include: 1. Training on AGD measurements for all examiners using well-described methods (preferred method is using calipers but other methods will be considered). 2. Intra- and inter-rater reliability assessed. 3. For hypospadias diagnosis, direct exam by urologists/pediatric urologist or examiners who participated in training. 4. Testosterone and/or free testosterone measured in serum or amniotic fluid by HPLC, GC-MS, LC-MS, or equilibrium dialysis.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), such as non-caliper AGD measurements or testosterone measurements using radioimmunoassay, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, 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, which is more likely to apply to objective outcome measures.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), or AGD assessment without intra-rater and/or inter-rater reliability, or hypospadias measured from the medical record, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument or no training for AGD measurement, or no description of AGD measurement or hypospadias assessment methods, or no description of methods for testosterone assays, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes 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 in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes 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 a composite score without individual outcome components or outcomes reported using measurements, analysis methods, or subsets of the data (e.g., subscales) that were not prespecified or reporting outcomes not prespecified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity? There are no phthalate-specific additions to the risk of bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk of bias considerations that do not fit under the other questions.

Cross-Sectional and Case-Series Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups? [NA to Case Series]

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and nonexposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates. • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed, but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and nonexposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and nonexposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group, including a different rate of nonresponse without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and nonexposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

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

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias, including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that 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, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as key/primary potential confounders and/or effect measure modifiers that must be considered in the analyses of the relationship between phthalate exposure and male reproductive outcomes: 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. • **Note:** The following variables should be considered as additional potential confounders and/or effect measure modifiers, but consideration of these variables is not required in the analysis of the relationship between phthalate exposure and male reproductive outcomes: maternal age, pre-pregnancy or maternal BMI, maternal education, maternal income, maternal race/ethnicity, and time of day of urine collection.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **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, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** This includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important 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 (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **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 (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using nonvalid measurements, • **OR** 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.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note:** Unacceptable handling of subject exclusion from analyses includes reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time frame) using well-established methods that directly measure exposure (e.g., measurement of urinary phthalate metabolites [and a measure of urinary dilution was available], amniotic fluid oxidative phthalate metabolites). • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the male reproductive outcome was assessed using well-established methods (e.g., gold standard), • **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **Note:** Well-established methods will depend on the outcome, and include: 1. Training on AGD measurements for all examiners using well-described methods (preferred method is using calipers but other methods will be considered). 2. Intra- and inter-rater reliability assessed. 3. For hypospadias diagnosis, direct exam by urologists/pediatric urologist or examiners who participated in training. 4. Testosterone and/or free testosterone measured in serum or amniotic fluid by HPLC, GC-MS, LC-MS, or equilibrium dialysis.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), such as non-caliper AGD measurements and testosterone measurements using radioimmunoassay,

• **AND** subjects had been followed for the same length of time in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, 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, which is more likely to apply to objective outcome measures.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), or AGD assessment without intra-rater and/or inter-rater reliability, or hypospadias measured from the medical record, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, or no training for AGD measurement, or no description of AGD measurement or hypospadias assessment methods, or no description of methods for testosterone assays, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes 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 in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes 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, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no phthalate-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Case-Control Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome. • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4)
Probably Low Risk of Bias (+)
• Indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age), recruited within the same time frame, and controls are described as having no history of the outcome, • **OR** it is deemed differences between cases and controls would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, • **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames.

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

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that 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, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as key/primary potential confounders and/or effect measure modifiers that must be considered in the analyses of the relationship between phthalate exposure and male reproductive outcomes: 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.

• **Note:** The following variables should be considered as additional potential confounders and/or effect measure modifiers but consideration of these variables is not required in the analysis of the relationship between phthalate exposure and male reproductive outcomes: maternal age, pre-pregnancy or maternal BMI, maternal education, maternal income, maternal race/ethnicity, and time of day of urine collection. • **Note:** It may be that in case control studies, the original cases and controls were matched on the covariates above. If this is the case, the adjustment is not needed.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **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, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** This includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important 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 (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **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 (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** 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.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note:** Unacceptable handling of subject exclusion from analyses includes reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of urinary phthalate metabolites[and a measure of urinary dilution was available], amniotic fluid oxidative phthalate metabolites). • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the male reproductive outcome was assessed using well-established methods (e.g., gold standard), • **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **Note:** Well-established methods will depend on the outcome, and include: 1. Training on AGD measurements for all examiners using well-described methods (preferred method is using calipers but other methods will be considered). 2. Intra- and inter-rater reliability assessed. 3. For hypospadias diagnosis, direct exam by urologists/pediatric urologist or examiners who participated in training. 4. Testosterone and/or free testosterone measured in serum or amniotic fluid by HPLC, GC-MS, LC-MS, or equilibrium dialysis.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), such as non-caliper AGD measurements and testosterone measurements using radioimmunoassays, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, 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, which is more likely to apply to objective outcome measures.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), or AGD assessment without intra-rater and/or inter-rater reliability, or hypospadias measured from the medical record, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, or no training for AGD measurement, or no description of AGD measurement or hypospadias assessment methods, or no description of methods for testosterone assays, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes 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 in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes 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 prespecified or reporting outcomes not prespecified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no phthalate-specific additions to the risk of bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk of bias considerations that do not fit under the other questions.

SECTION D-1f

AMENDMENTS TO THE PROTOCOL

Additions to the Review Team

The following committee member was added to the review team supplement expertise and to assist with the workload:

- **David C. Dorman** (*Chair*) is a professor of toxicology in the Department of Molecular Biosciences of North Carolina State University. The primary objective of his research is to provide a refined understanding of chemically induced neurotoxicity in laboratory animals that will lead to improved assessment of potential toxicity in humans. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has chaired or served on several NRC committees, including the Committee on Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decisions, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He has served on other advisory boards for the US Navy, NASA, and USDA and is currently a member of NTP's Board of Scientific Counselors. Dr. Dorman is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Sciences. He received a DVM from Colorado State University. He completed a combined PhD and residency program in toxicology at the University of Illinois at Urbana-Champaign, and he is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.

A consulting firm (ICF International) was hired to assist the committee with extracting data from the epidemiological studies into HAWC. The two ICF staff members who performed the extraction task were:

- **Robyn Blain**, who has 22 years of experience reviewing and analyzing public-health and mammalian toxicity studies, with 14 years at ICF. She also has about 10 years of experience reviewing and analyzing epidemiologic and mechanism studies. She has applied her expertise in several work assignments for NTP, using both DRAGON and HAWC as well as the Excel-based ROBINS-E risk of bias tool. Dr. Blain also has been involved in several work assignments for the US Environmental Protection Agency's (EPA's) National Center for Environmental Assessment (NCEA), including authoring Provisional Toxicity Values (PTVs) in support of EPA's Superfund Program and several Integrated Risk Information System (IRIS) Toxicological Reviews; supporting a 2,3,7,8-tetrachlorodibenzo-p-dioxin literature review; evaluating studies for chemical registration under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Program; developing robust study summaries in International Uniform Chemical Information Database (IUCLID) version 5; and preparing toxicological assessments for several chemicals. She has conducted multiple literature reviews and has been involved in the production and testing of two relational retrieval databases. Dr. Blain has extensive experience with in vivo toxicological experimentation and has published several articles on the hepatotoxicity of organic solvents, the induction of cancer from single exposures to carcinogens, and hormesis.
- **Pamela Hartman**, who has more than 20 years of professional experience in environmental consulting, specializing in exposure and risk assessment, toxicology, literature search and review, technical editing, and document production. For NIEHS, she has conducted data extraction, study quality reviews, and risk of bias assessments for toxicological and epidemiological studies using DRAGON and HAWC for multiple projects, including perfluorooctanoic acid (PFOA)/perfluorooctane sulfonate (PFOS), bisphenol A (BPA), Fluoride, Folic Acid, and Transgenerational

Inheritance. Ms. Hartman has also provided support to many work assignments for EPA/NCEA, specifically: Exposure Factors Interactive Resource for Scenarios Tool (ExpoFIRST); numerous IRIS Toxicological Reviews; EPA-Expo-Box; HERO Support; Risk Assessment Training and Experience (RATE) Program—Exposure Assessment (EXA) Course Series; Provisional Toxicity Value (PTV) documents; two Nanomaterial Case Study documents; and Dioxin Reassessment. Ms. Hartman has a BS in Natural Resources from Cornell University and an MA in Environmental Management from Duke University.

SECTION D-2

Results of Literature Searches for Human Studies on the Effects of Phthalates on Male Reproductive-Tract Development

Literature searches were performed on August 15, 2016, using the search strategy presented in the Phthalate (Human) Systematic Review Protocol (Section D-1). A summary of the results is presented below.

Embase:	422
PubMed:	210
Toxline:	111

Total citations found:	743
Duplicates removed:	149
Total unique citations:	594

SECTION D-3

Confidence Ratings for the Body of Evidence from Human Studies of Phthalates and Anogenital Distance

The confidence in the body of evidence from human studies on phthalates and male reproductive-tract development was rated in accordance with the OHAT Guidance (NTP 2015) specified in Section D-1. The results for DEHP are presented first, and the remaining phthalates are subsequently presented in alphabetical order.

DEHP (metabolites MEHP, 5oxo-MEHP, 5OH-MEHP, or sumDEHP Metabolites)

Five human studies of DEHP and AGD (as) or AGD (ap) were available. Figures D3-1 and D3-2 illustrate the data from studies that evaluated sumDEHP metabolites and individual DEHP metabolites, respectively.

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous sum DEHP metabolites (log transformed)	sum DEHP metabolites
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous sum DEHP metabolites (log transformed)	sum DEHP metabolites
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 2)	167	continuous sum DEHP T2 (log10 SG adjusted)	sum DEHP T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 3)	167	continuous sum DEHP T3 (log10 SG adjusted)	sum DEHP T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 2)	168	continuous sum DEHP T2 (log10 SG adjusted)	sum DEHP T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 3)	168	continuous sum DEHP T3 (log10 SG adjusted)	sum DEHP T3
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous sum DEHP metabolites (Log-transformed)	sum DEHP metabolites
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous sum DEHP metabolites (Log-transformed)	sum DEHP metabolites
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous sum DEHP metabolites (log transformed)	sum DEHP metabolites
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous sum DEHP metabolites (log transformed)	sum DEHP metabolites

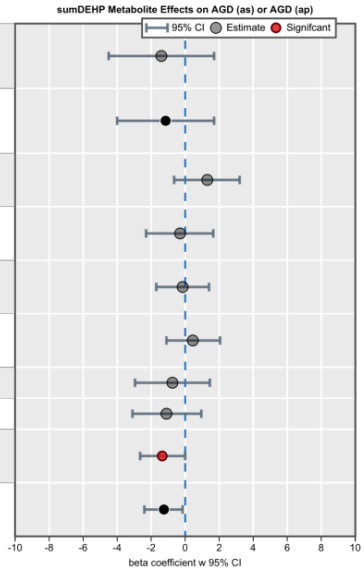
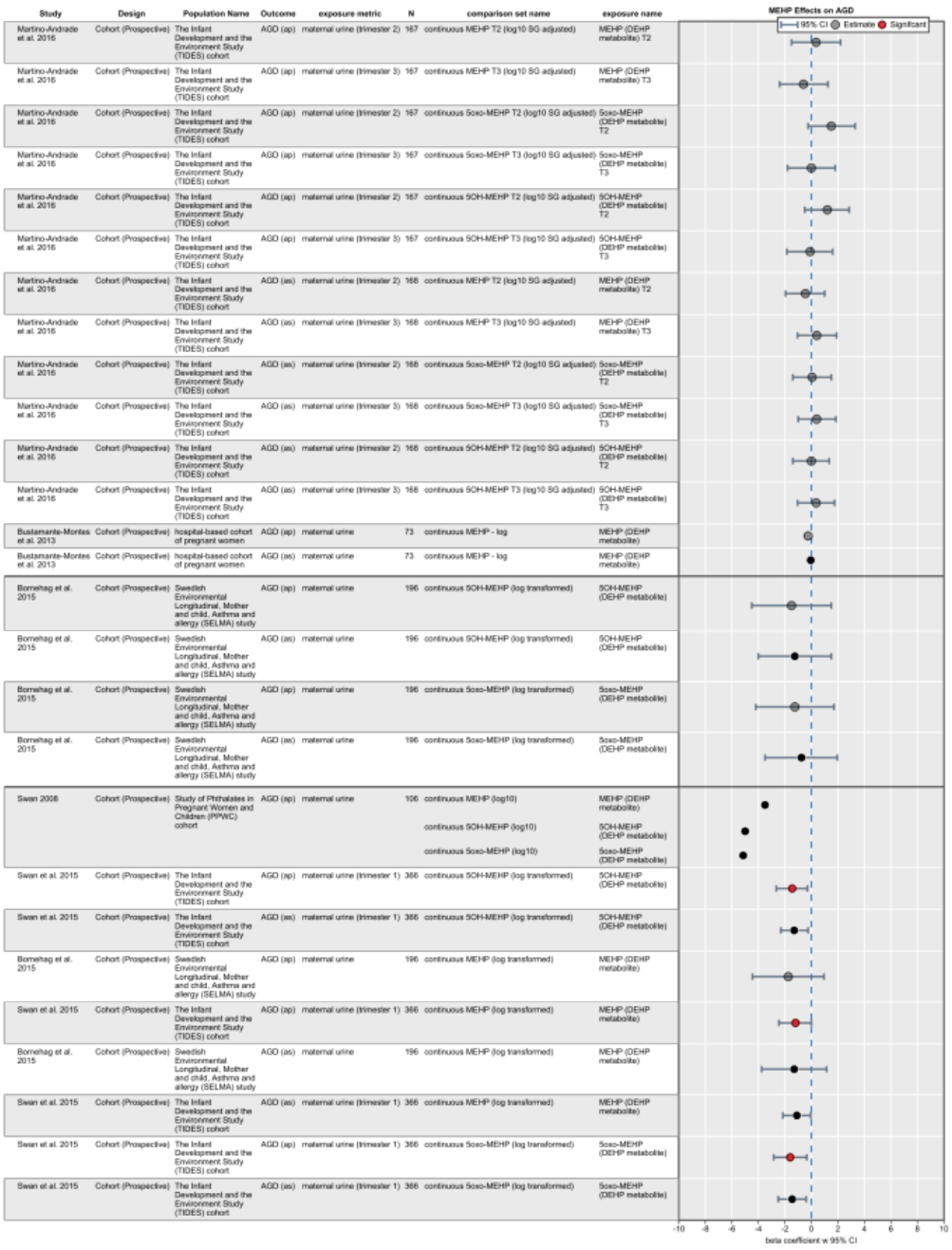


FIGURE D3-1 Data pivot of studies that measured sumDEHP metabolites and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/sumdehp-metabolite-effects-agd-or-agd-ap/.



Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 2)	167	continuous MEHP T2 (log10 SG adjusted)	MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 3)	167	continuous MEHP T3 (log10 SG adjusted)	MEHP (DEHP metabolite) T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 2)	167	continuous 5oxo-MEHP T2 (log10 SG adjusted)	5oxo-MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 3)	167	continuous 5oxo-MEHP T3 (log10 SG adjusted)	5oxo-MEHP (DEHP metabolite) T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 2)	167	continuous 5OH-MEHP T2 (log10 SG adjusted)	5OH-MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 3)	167	continuous 5OH-MEHP T3 (log10 SG adjusted)	5OH-MEHP (DEHP metabolite) T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 2)	168	continuous MEHP T2 (log10 SG adjusted)	MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 3)	168	continuous MEHP T3 (log10 SG adjusted)	MEHP (DEHP metabolite) T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 2)	168	continuous 5oxo-MEHP T2 (log10 SG adjusted)	5oxo-MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 3)	168	continuous 5oxo-MEHP T3 (log10 SG adjusted)	5oxo-MEHP (DEHP metabolite) T3
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 2)	168	continuous 5OH-MEHP T2 (log10 SG adjusted)	5OH-MEHP (DEHP metabolite) T2
Martino-Andrade et al. 2016	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 3)	168	continuous 5OH-MEHP T3 (log10 SG adjusted)	5OH-MEHP (DEHP metabolite) T3
Bustamante-Montes et al. 2013	Cohort (Prospective)	hospital-based cohort of pregnant women	AGD (ap)	maternal urine	73	continuous MEHP - log	MEHP (DEHP metabolite)
Bustamante-Montes et al. 2013	Cohort (Prospective)	hospital-based cohort of pregnant women	AGD (as)	maternal urine	73	continuous MEHP - log	MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous 5OH-MEHP (log transformed)	5OH-MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous 5OH-MEHP (log transformed)	5OH-MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous 5oxo-MEHP (log transformed)	5oxo-MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous 5oxo-MEHP (log transformed)	5oxo-MEHP (DEHP metabolite)
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MEHP (log10)	MEHP (DEHP metabolite)
						continuous 5OH-MEHP (log10)	5OH-MEHP (DEHP metabolite)
						continuous 5oxo-MEHP (log10)	5oxo-MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous 5OH-MEHP (log transformed)	5OH-MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous 5OH-MEHP (log transformed)	5OH-MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous MEHP (log transformed)	MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MEHP (log transformed)	MEHP (DEHP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous MEHP (log transformed)	MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MEHP (log transformed)	MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous 5oxo-MEHP (log transformed)	5oxo-MEHP (DEHP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous 5oxo-MEHP (log transformed)	5oxo-MEHP (DEHP metabolite)

FIGURE D3-2 Data pivot of studies that measured DEHP metabolites and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mehp-effects-agd-and-agd-ap/.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. Figure D3-3 shows the risk of bias assessment for each study.
- **Unexplained inconsistencies:** No downgrade. Although there is some inconsistency in results across studies, the differences can be explained (at least partially) by different study populations, different ranges of exposure within study population, and differences in timing of collection of urine samples (first trimester vs later samples). Despite these differences, there is general consistency of associations with respect to direction and magnitude of effects (smaller AGD with higher DEHP metabolite exposure).
- **Indirectness:** No downgrade. The studies directly addressed the effect of prenatal exposure to DEHP on AGD in males, defined the window of exposure, and assessed the outcome within an appropriate amount of time.
- **Imprecision:** No downgrade. Most of the studies included 95% confidence intervals to assess precision of associations.
- **Publication bias:** No downgrade (see Table D3-1).

Sources of funding were used to evaluate publication bias in terms of whether a particular sector funded more studies than another.

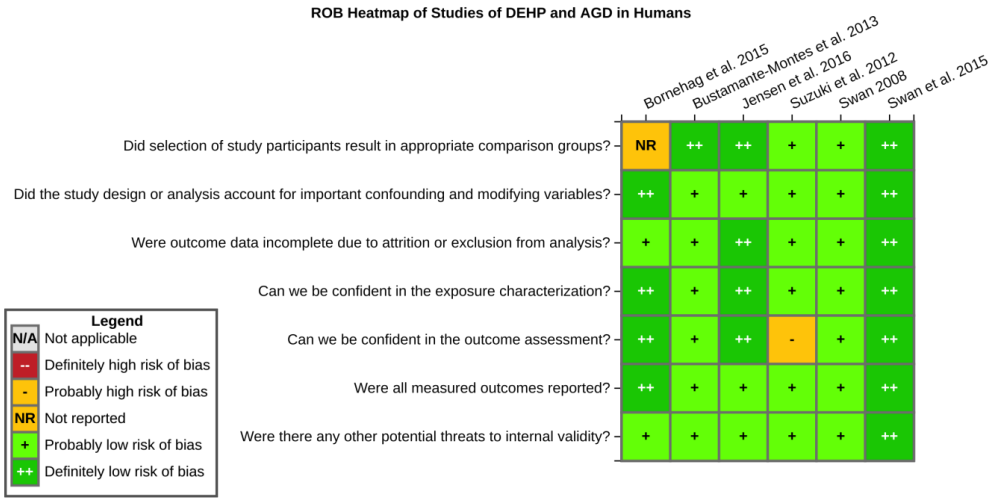


FIGURE D3-3 Risk of bias heatmap of studies of DEHP and AGD in humans. In HAWC: https://hawcproject.org/summary/visual/341/.

TABLE D3-1 Sources of Funding for the Human Studies on Phthalates

Reference	Government	Industry	Other	Unknown
Bornehag et al. 2015	X (Sweden)			
Bustamante-Montes et al. 2013	X (Mexico)			
Jensen et al. 2016	X (Denmark)		X (The Ronald McDonald Children Foundation, K.A. Rhode Foundation, The Danielsen Foundation)	
Suzuki et al. 2012			X (Japan Society for the Promotion of Science)	
Swan 2008	X (EPA, NIH, State of Iowa)*			
Swan et al. 2015/Martino-Andrade et al. 2015	X (NIEHS)			

*Funding information obtained from Swan et al. (2005).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. No evidence of a large magnitude of effect size.
- **Dose-response:** No upgrade. No evidence of dose-response curve based on one study that performed quartile analyses (see Figure D3-4).
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

BzBP (metabolite MBzP)

Four human studies of BzBP and AGD (as) or AGD (ap) were available (see Figure D3-5).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure D3-3 for the risk of bias assessments for the four studies.
- **Unexplained inconsistencies:** No downgrade. Although there is some inconsistency in results across studies, the differences can be explained (at least partially) by different study populations, different ranges of exposure within the study population, and differences in timing of collection of urine samples (first trimester vs later samples).

Study	Design	Population Name	Outcome	exposure metric	N	exposure group name	exposure name
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine		1st quartile (LOD-13.9 ng/mL)	sum DEHP metabolites
						2nd quartile (14.0-21.9 ng/mL)	sum DEHP metabolites
						3rd quartile (22.0-33.9 ng/mL)	sum DEHP metabolites
						4th quartile (>=34 ng/mL)	sum DEHP metabolites
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine		1st quartile (LOD-13.9 ng/mL)	sum DEHP metabolites
						2nd quartile (14.0-21.9 ng/mL)	sum DEHP metabolites
						3rd quartile (22.0-33.9 ng/mL)	sum DEHP metabolites
						4th quartile (>=34 ng/mL)	sum DEHP metabolites

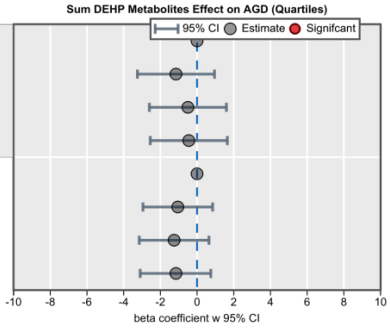


FIGURE D3-4 Data pivot of the Jensen et al. (2016) study of sumDEHP metabolites and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/sum-dehp-effects-agd-ap-agd-and-agi-quartiles/.

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MBzP (log10)	MBzP (BBzP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous MBzP (log transformed)	MBzP (BBzP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous MBzP (Log-transformed)	MBzP (BBzP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MBzP (log transformed)	MBzP (BBzP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous MBzP (log transformed)	MBzP (BBzP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous MBzP (Log-transformed)	MBzP (BBzP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MBzP (log transformed)	MBzP (BBzP metabolite)

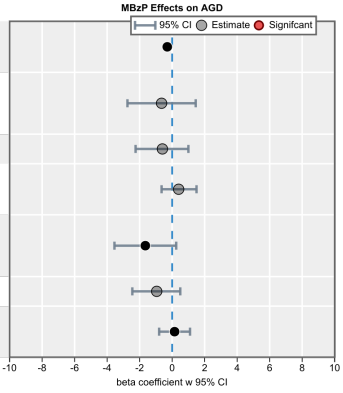


FIGURE D3-5 Data pivot of studies that measured MBzP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mbzp-effects-agd-ap-or-agd/.

- **Indirectness:** No downgrade. The study designs directly addressed the topic of the evaluation.
- **Imprecision:** No downgrade. All but one study (Swan 2008) included 95% confidence intervals to assess precision of associations.
- **Publication bias:** No downgrade (see Table D3-1).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. No evidence of a large magnitude of effect size.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

DBP (metabolite MBP)

Four human studies of MBP and AGD (as) or AGD (ap) were available (see Figure D3-6).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure D3-3 for the risk of bias assessments for the four studies.
- **Unexplained inconsistencies:** No downgrade. Although there is some inconsistency in results across studies, the differences can be explained (at least partially) by different study populations, different ranges of exposure within the study population, and differences in timing of collection of urine samples (first trimester vs later samples). Despite these differences, there is general consistency of associations with respect to direction and magnitude of effects (smaller AGD with higher MBP metabolite exposure).
- **Indirectness:** No downgrade. The study designs directly addressed the topic of the evaluation.
- **Imprecision:** No downgrade. All but one study (Swan 2008) included 95% confidence intervals to assess precision of associations.
- **Publication bias:** No downgrade (see Table D3-1).

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MBP (log10)	MBP (DBP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous MBP (log transformed)	MBP (DBP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous MBP (Log-transformed)	MBP (DBP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous MBP (log transformed)	MBP (DBP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous MBP (Log-transformed)	MBP (DBP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	305	continuous sum MBP (Log-transformed)	sum MBP
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous sum MBP (Log-transformed)	sum MBP
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MBP (log transformed)	MnBP (DBP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MBP (log transformed)	MnBP (DBP metabolite)

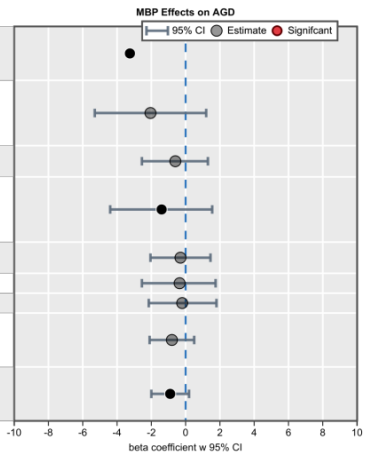


FIGURE D3-6 Data pivot of studies that measured MBP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mbp-effects-agd-or-agd-ap/update/.

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. No evidence of a large magnitude of effect size.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

DEP (metabolite MEP)

Four human studies of DEP and AGD (as) or AGD (ap) were available (see Figure D3-7).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure D3-3 for the risk of bias assessments for the four studies.
- **Unexplained inconsistencies:** No downgrade. Results are largely null except for Swan (2008).
- **Indirectness:** No downgrade. The study designs directly addressed the topic of the evaluation.
- **Imprecision:** No downgrade. All but one study (Swan 2008) included 95% confidence intervals to assess precision of associations.
- **Publication bias:** No downgrade (see Table D3-1).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. No evidence of a large magnitude of effect size.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MEP (log10)	MEP (DEP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous MEP (log transformed)	MEP (DEP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous MEP (Log-transformed)	MEP (DEP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MEP (log transformed)	MEP (DEP metabolite)
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous MEP (log transformed)	MEP (DEP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous MEP (Log-transformed)	MEP (DEP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MEP (log transformed)	MEP (DEP metabolite)

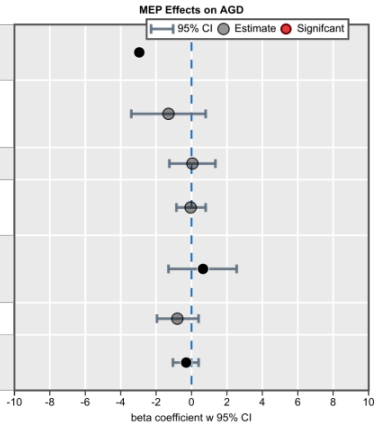


FIGURE D3-7 Data pivot of studies that measured MEP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mep-effects-agd-or-agd-ap/.

DIBP (metabolite MIBP)

Three human studies of DIBP and AGD (as) or AGD (ap) were available (see Figure D3-8).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure D3-3 for the risk of bias assessments for the three studies.
- **Unexplained inconsistencies:** No downgrade. Results are largely null, except for Swan (2008), and they are consistent across two other larger studies (Swan et al. 2015; Jensen et al. 2016).
- **Indirectness:** No downgrade. The study designs directly addressed the topic of the evaluation.
- **Imprecision:** No downgrade. All but one study (Swan 2008) included 95% confidence intervals to assess precision of associations.
- **Publication bias:** No downgrade (see Table D3-1).

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MiBP (log10)	MiBP (DiBP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous MiBP (Log-transformed)	MiBP (DiBP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MiBP (log transformed)	MiBP (DiBP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous MiBP (Log-transformed)	MiBP (DiBP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MiBP (log transformed)	MiBP (DiBP metabolite)

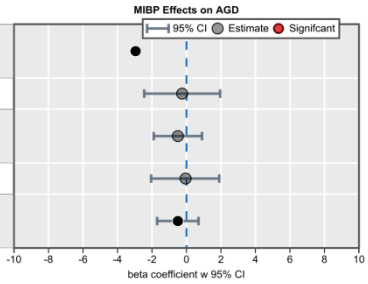


FIGURE D3-8 Data pivot of studies that measured MIBP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mibp-effects-agd-or-agd-ap/.

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. The effect estimates were generally null.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

DIDP (metabolite MCNP)

One human study that evaluated the relationship between metabolites of DIDP and AGD was available (see Figure D3-9).

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)		continuous MCNP (log transformed)	MCNP (DIDP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)		continuous MCNP (log transformed)	MCNP (DIDP metabolite)

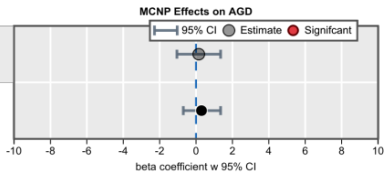


FIGURE D3-9 Data pivot of the study that measured MCNP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mcnp-effects-agd-or-agd-ap/.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. The study was rated as having a definitely low risk of bias (see Figure D3-3).
- **Unexplained inconsistencies:** No downgrade.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade.
- **Publication bias:** No downgrade (see Table D3-1).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. The effect estimates were generally null.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Study measured and adjusted for important known confounders.

DINP (metabolite MCOP)

Three human studies that evaluated the relationship between metabolites of DINP and AGD were available (see Figure D3-10).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. See Figure D3-3 for the risk of bias assessments for the three studies.
- **Unexplained inconsistencies:** No downgrade. Although there is some inconsistency in results across studies, the differences can be explained (at least partially) by different study populations, different ranges of exposure within study population, and differences in timing of collection of urine samples (first trimester vs later samples).
- **Indirectness:** No downgrade. The study designs directly addressed the topic of the evaluation.
- **Imprecision:** No downgrade.
- **Publication bias:** No downgrade (see Table D3-1).

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (ap)	maternal urine	196	continuous sum DiNP metabolites (log transformed)	sum DiNP metabolites
Bornehag et al. 2015	Cohort (Prospective)	Swedish Environmental Longitudinal, Mother and child, Asthma and allergy (SELMA) study	AGD (as)	maternal urine	196	continuous sum DiNP metabolites (log transformed)	sum DiNP metabolites
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)		continuous MCOP (DINP metabolite) (log transformed)	MCOP (DINP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)		continuous MCOP (DINP metabolite) (log transformed)	MCOP (DINP metabolite)
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (ap)	maternal urine	236	continuous sum DiNP metabolites (Log-transformed)	sum DiNP metabolites
Jensen et al. 2016	Cohort (Prospective)	Odense child cohort	AGD (as)	maternal urine	245	continuous sum DiNP metabolites (Log-transformed)	sum DiNP metabolites

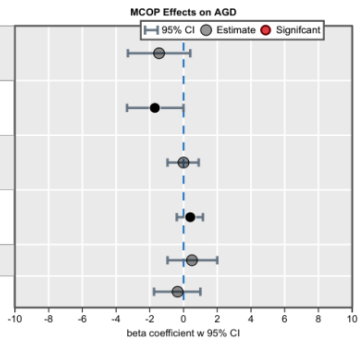


FIGURE D3-10 Data pivot of studies that measured MCOP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mcop-effects-agd-or-agd-ap/.

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. No evidence of a large magnitude of effect size.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders. Little evidence that unmeasured confounding would affect the results across studies. Direction of potential residual confounding bias is unknown.

DMP (metabolite MMP)

One human study of the relationship between metabolites of DMP and AGD was available (see Figure D3-11).

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MMP (log10)	MMP (DMP metabolite)

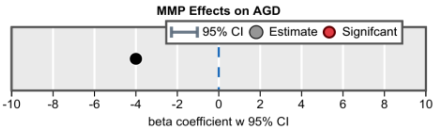


FIGURE D3-11 Data pivot of the study that measured MMP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mmp-effects-agd-or-agd-ap/.

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. The study was rated as having a probably low risk of bias (see Figure D3-3).
- **Unexplained inconsistencies:** No downgrade.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade.
- **Publication bias:** No downgrade (see Table D3-1).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Study measured and adjusted for important known confounders.

DOP (metabolite MCPP)

Two human studies of the relationship between metabolites of DOP and AGD were available (see Figure D3-12).

Factors Considered for Downgrading Confidence

- **Risk of bias:** No downgrade. The studies were rated as having probably or definitely low risk of bias (see Figure D3-3).
- **Unexplained inconsistencies:** No downgrade.
- **Indirectness:** No downgrade.
- **Imprecision:** No downgrade.

- **Publication bias:** No downgrade (see Table D3-1).

Factors Considered for Upgrading Confidence

- **Large magnitude:** No upgrade. The effect estimates were generally null.
- **Dose-response:** No upgrade. Unable to assess because no quartile data were available.
- **Residual confounding:** No upgrade. Studies measured and adjusted for important known confounders.

Study	Design	Population Name	Outcome	exposure metric	N	comparison set name	exposure name
Swan 2008	Cohort (Prospective)	Study of Phthalates in Pregnant Women and Children (PPWC) cohort	AGD (ap)	maternal urine	106	continuous MCPP (log10)	MCPP (DOP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (ap)	maternal urine (trimester 1)	366	continuous MCPP (log transformed)	MCPP (DOP metabolite)
Swan et al. 2015	Cohort (Prospective)	The Infant Development and the Environment Study (TIDES) cohort	AGD (as)	maternal urine (trimester 1)	366	continuous MCPP (log transformed)	MCPP (DOP metabolite)

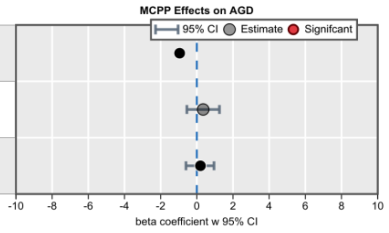


FIGURE D3-12 Data pivot of studies that measured MCPP and AGD (as) or AGD (ap). In HAWC: https://hawcproject.org/summary/data-pivot/assessment/350/mcpp-effects-agd-or-agd-ap/.

SECTION D-4

Sensitivity Analyses of DEHP and AGD

TABLE D4-1 Sensitivity Analyses Performed by Leaving One Study Out at a Time, Using Alternative Exposure and Outcome Measures for Each Study One at a Time, and Restricting Analyses to Use the Same Exposure Measure (sumDEHP or MEHP) and/or the Same Outcome Measure (AGD [as] or AGD [ap])

Analysis	Estimate, mm	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
Primary Analysis	-4.07	-6.49	-1.66	0.001	0.876	0.0
w/o Bornehag et al. 2015	-4.25	-6.83	-1.68	0.001	0.786	0.0
w/o Bustamante-Montes et al. 2013	-4.31	-6.79	-1.82	0.001	0.886	0.0
w/o Jensen et al. 2016.0	-4.35	-7.03	-1.66	0.002	0.800	0.0
w/o Swan 2008.0	-3.59	-6.58	-0.60	0.019	0.820	0.0
w/o Swan et al. 2015.0	-3.68	-6.52	-0.85	0.011	0.814	0.0
Using Using Alternate Estimates						
Bornehag et al. 2015.1	-4.07	-6.43	-1.71	0.001	0.881	0.0
Bornehag et al. 2015.2	-3.93	-6.33	-1.53	0.001	0.826	0.0
Bornehag et al. 2015.3	-4.09	-6.49	-1.69	0.001	0.881	0.0
Bornehag et al. 2015.4	-3.99	-6.39	-1.58	0.001	0.847	0.0
Bornehag et al. 2015.5	-3.44	-5.56	-1.31	0.002	0.686	0.0
Bornehag et al. 2015.6	-3.44	-5.46	-1.41	0.001	0.723	0.0
Bornehag et al. 2015.7	-3.32	-5.41	-1.23	0.002	0.639	0.0
Bornehag et al. 2015.8	-3.44	-5.55	-1.34	0.001	0.699	0.0
Bornehag et al. 2015.9	-3.13	-5.37	-0.90	0.006	0.503	8.4
Bustamante-Montes et al. 2013	-3.46	-5.66	-1.27	0.002	0.620	0.0
Jensen et al. 2016.1	-3.06	-5.36	-0.76	0.009	0.494	14.5
Swan 2008.1	-4.78	-7.24	-2.32	0.000	0.590	0.0
Swan 2008.2	-4.66	-7.15	-2.17	0.000	0.640	0.0
Martino-Andrade et al. 2016.1	-3.16	-5.74	-0.58	0.017	0.787	0.0
Martino-Andrade et al. 2016.2	-3.35	-5.91	-0.79	0.010	0.872	0.0
Martino-Andrade et al. 2016.3	-2.93	-5.48	-0.39	0.024	0.673	0.0
Martino-Andrade et al. 2016.4	-2.92	-5.44	-0.41	0.023	0.689	0.0
Martino-Andrade et al. 2016.5	-3.20	-5.79	-0.61	0.015	0.804	0.0
Martino-Andrade et al. 2016.6	-1.68	-5.07	1.70	0.330	0.113	47.6
Martino-Andrade et al. 2016.7	-2.16	-4.94	0.61	0.127	0.367	26.8
Martino-Andrade et al. 2016.8	-1.51	-5.08	2.06	0.406	0.055	54.3
Martino-Andrade et al. 2016.9	-1.64	-5.01	1.74	0.341	0.086	50.3

Martino-Andrade et al. 2016.10	-1.79	-5.05	1.48	0.283	0.156	43.4
Martino-Andrade et al. 2016.11	-2.69	-5.39	0.00	0.050	0.493	5.7
Martino-Andrade et al. 2016.12	-2.60	-5.33	0.13	0.062	0.484	9.0
Martino-Andrade et al. 2016.13	-2.53	-5.30	0.24	0.073	0.463	11.9
Martino-Andrade et al. 2016.14	-2.55	-5.28	0.18	0.067	0.481	10.8
Martino-Andrade et al. 2016.15	-2.91	-5.51	-0.32	0.028	0.612	0.0
Martino-Andrade et al. 2016.16	-2.66	-4.97	-0.35	0.024	0.658	0.0
Martino-Andrade et al. 2016.17	-2.76	-5.02	-0.51	0.016	0.727	0.0
Martino-Andrade et al. 2016.18	-2.35	-4.89	0.19	0.070	0.498	16.7
Martino-Andrade et al. 2016.19	-2.38	-4.86	0.09	0.059	0.527	14.5
Martino-Andrade et al. 2016.20	-2.82	-5.14	-0.51	0.017	0.734	0.0
Swan et al. 2015.21	-3.95	-6.30	-1.59	0.001	0.901	0.0
Swan et al. 2015.22	-4.33	-6.69	-1.97	0.000	0.810	0.0
Swan et al. 2015.23	-4.17	-6.50	-1.84	0.000	0.862	0.0
Swan et al. 2015.24	-3.75	-6.17	-1.33	0.002	0.916	0.0
Swan et al. 2015.25	-3.17	-5.12	-1.23	0.001	0.881	0.0
Swan et al. 2015.26	-2.97	-4.81	-1.12	0.002	0.848	0.0
Swan et al. 2015.27	-3.42	-5.30	-1.55	0.000	0.909	0.0
Swan et al. 2015.28	-3.26	-5.08	-1.44	0.000	0.895	0.0
Swan et al. 2015.29	-2.73	-4.69	-0.78	0.006	0.778	0.0
Additional Analysis						
Only sumDEHP	-3.91	-7.04	-0.78	0.014	0.787	0.0
Only AGD (as)	-3.59	-6.58	-0.60	0.019	0.820	0.0
Only MEHP	-4.17	-6.71	-1.62	0.001	0.833	0.0
Only AGD (ap)	-2.23	-3.78	-0.68	0.005	0.560	0.0
Only AGD (as) and sumDEHP	-3.91	-7.04	-0.78	0.014	0.787	0.0
Only AGD (ap) and sumDEHP	-1.96	-3.75	-0.17	0.032	0.730	0.0
Only AGD (as) and MEHP	-3.65	-6.90	-0.40	0.028	0.734	0.0
Only AGD (ap) and MEHP	-2.53	-4.18	-0.88	0.003	0.536	0.0

SECTION D-5

Meta-Analyses of Human Studies of Additional Phthalates and Anogenital Distance

Meta-analyses of human studies on BzBP, DBP, DEP, DIBP, and DINP in relation to alterations in anogenital distance (AGD) were conducted. The same meta-analysis methods used for DEHP in Chapter 3 were applied to these phthalate. (Three phthalates, DIDP, DMP, and DOP, had only one study each so no meta-analyses for these phthalates were performed.)

For each study, AGD (as) is preferred over AGD (ap). For the studies by Bustamonte-Montes et al. (2013) and Swan (2008), the confidence interval was estimated using the reported p-value, assuming a normal distribution.

Beta coefficients are reported in units of mm/$\log_{10}$ change in exposure. Two factors a priori may affect comparability across studies. First, there are baseline differences in AGD (as) across different studies due to demographic factors, such as birth weight. For instance, the mean AGD (as) in Bustamante-Montes et al. (2013) was 12.4 mm, whereas the mean AGD (as) in Bornehag et al. (2015) was 41.4 mm. Additionally, AGD (as) is shorter than AGD (ap) is. For instance, in the study by Jensen et al. (2016), mean AGD (as) was 36.9 mm whereas mean AGD (ap) was 70.2 mm. Therefore, the same mm change may reflect different percentage change in AGD across studies in end points. 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 is standardized to a percent change in AGD per $\log_{10}$ change in exposure.

Sensitivity analyses included leaving one study out at a time and using AGD (ap) exclusively as the outcome measure. As separate meta-analysis for using exclusive AGD (as) was not performed because it is the same as excluding the Swan (2008) study.

BzBP Meta-Analysis

Primary Analysis

In the primary analysis, four studies (see Table D5-1), with beta coefficients standardized to a percent change per $\log_{10}$ change in BzBP exposure, were analyzed using a random effects model. A summary estimate of -1.43 [95% CI: -3.47, 0.61] ($p = 0.17$) was found (see Figure D5-1). There was no significant heterogeneity, with an estimated I^2 value of 0% (Q statistic was not statistically significant). In the sensitivity analyses (see Figures D5-2 and D5-3 and Table D5-2), effect sizes ranged from -0.15 to -2.21, none of which were statistically significant. In sum, although a small effect was observed, the precision of the estimate was not sufficient to rule out chance. Thus, the available studies do not support BzBP exposure being associated with decreased AGD.

TABLE D5-1 Studies Included in the Meta-Analysis of BzBP and AGD

Reference	Outcome	Mean AGD, mm	Exposure Metric	Estimate, %	Lower CI	Upper CI
Bornehag et al. 2015	AGD (as)	41.40	MBzP (maternal urine)	-4.01	-8.60	0.60
Jensen et al. 2016	AGD (as)	36.90	MBzP (maternal urine)	-2.63	-6.61	1.30
Swan et al. 2015	AGD (as)	24.73	MBzP (maternal urine, trimester 1)	0.65	-3.28	4.53
Swan 2008	AGD (ap)	70.40	MBzP (maternal urine)	-0.44	-4.41	3.52

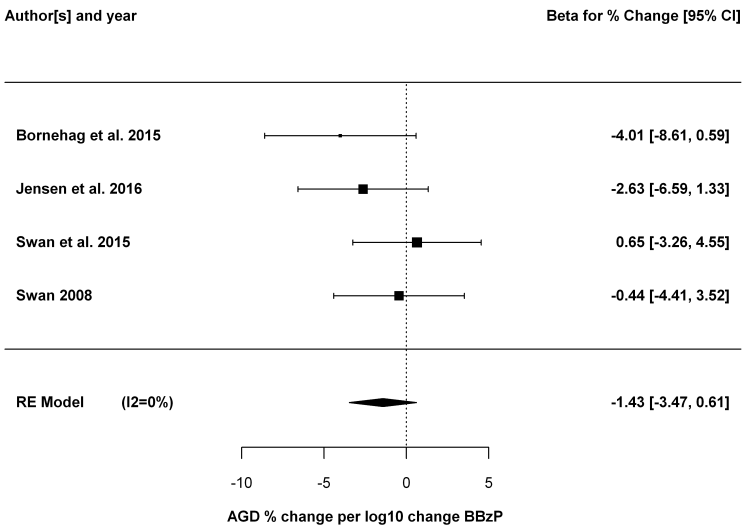


FIGURE D5-1 Meta-analysis of human studies of BzBP and AGD; reported effect estimates [95% confidence interval] from individual studies and overall pooled estimate from random effects (RE) model per 10-fold increase in BzBP exposure.

Sensitivity Analyses

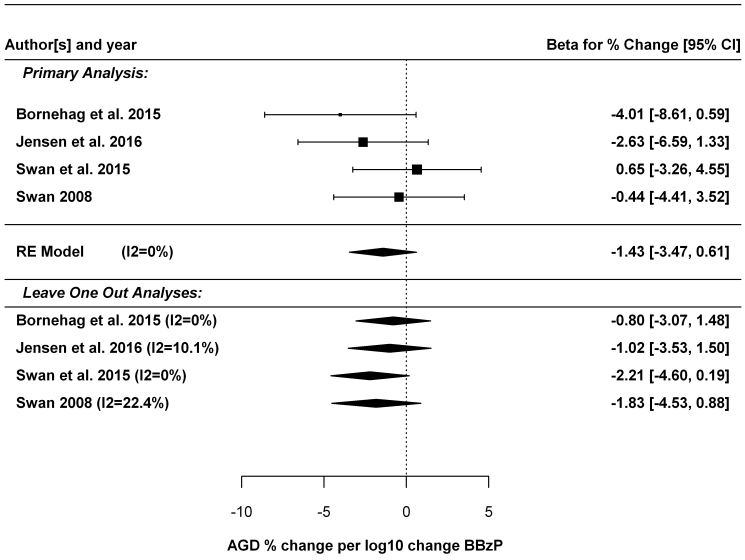


FIGURE D5-2 Sensitivity analyses of human studies of BzBP and AGD performed by leaving one study out at a time.

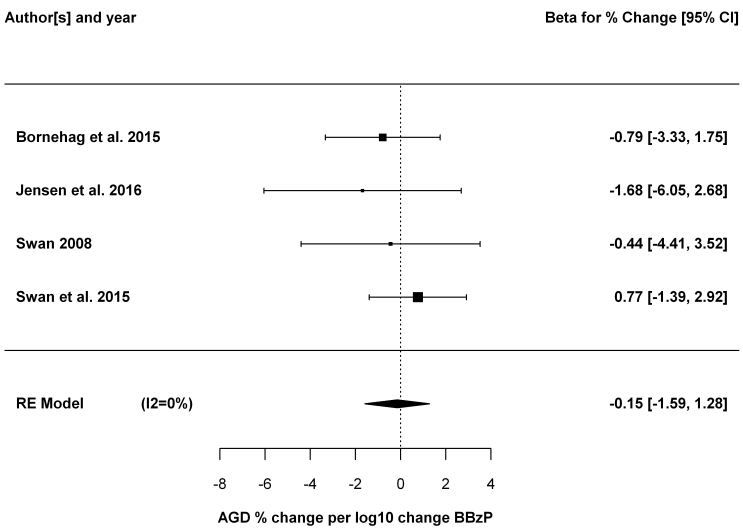


FIGURE D5-3 Sensitivity analyses of human studies of BzBP and AGD performed by restricting analysis to the same outcome measure (AGD [ap]).

TABLE D5-2 Sensitivity Analyses of Human Studies of BzBP and AGD

Analysis	Beta Estimate, % change	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
BzBP primary analysis	-1.43	-3.47	0.61	0.170	0.410	0.0
BzBP w/o Bornehag et al. 2015	-0.80	-3.07	1.48	0.493	0.502	0.0
BzBP w/o Jensen et al. 2016	-1.02	-3.53	1.50	0.428	0.301	10.1
BzBP w/o Swan et al. 2015	-2.21	-4.60	0.19	0.071	0.498	0.0
BzBP w/o Swan 2008	-1.83	-4.53	0.88	0.186	0.277	22.4
BzBP only AGD (ap)	-0.15	-1.59	1.28	0.835	0.698	0.0

DBP Meta-Analysis

Primary Analysis

In the primary analysis, four studies (see Table D5-3), with beta coefficients standardized to a percent change per $\log_{10}$ change in DBP exposure, were analyzed using a random effects model. A summery estimate of -3.13 [95% CI: -5.63, -0.64] (p = 0.014) was found (see Figure D5-4). There was no significant heterogeneity, with an estimated I^2 value of 0% (Q statistic was not statistically significant). In the sensitivity analyses (see Figures D5-5 and D5-6 and Table D5-4), effect sizes ranged from -1.85 to -4.02, and remained statistically significant in three of the five analyses. Specifically, dropping either the Swan (2008) or Swan et al. (2015) studies resulted in summary estimates that were no longer statistically significant. There was no observed heterogeneity in any sensitivity analysis results ($I^2 = 0$).

Overall, there is consistent evidence of a small decrease in AGD being associated with increasing DBP exposure, of magnitude around 3% for each $\log_{10}$ increase in DBP exposure. However, some uncertainty remains because the statistical significance of this result depends on the Swan (2008) or Swan et al. (2015) studies. On the other hand, there was no observed heterogeneity, so it is likely that this sensitivity is related to the decreased statistical power when dropping studies.

TABLE D5-3 Studies Included in the Meta-Analysis of DBP and AGD

Reference	Outcome	Mean AGD, mm	Exposure Metric	Estimate, %	Lower CI	Upper CI
Bornehag et al. 2015	AGD (as)	41.40	MBP (maternal urine)	-3.41	-10.60	3.79
Swan et al. 2015	AGD (as)	24.73	MBP (maternal urine, trimester 1)	-3.68	-8.17	0.81
Jensen et al. 2016	AGD (as)	36.90	MBP (maternal urine)	-0.81	-5.56	3.93
Swan 2008	AGD (ap)	70.40	MBP (maternal urine)	-4.62	-9.23	-0.02

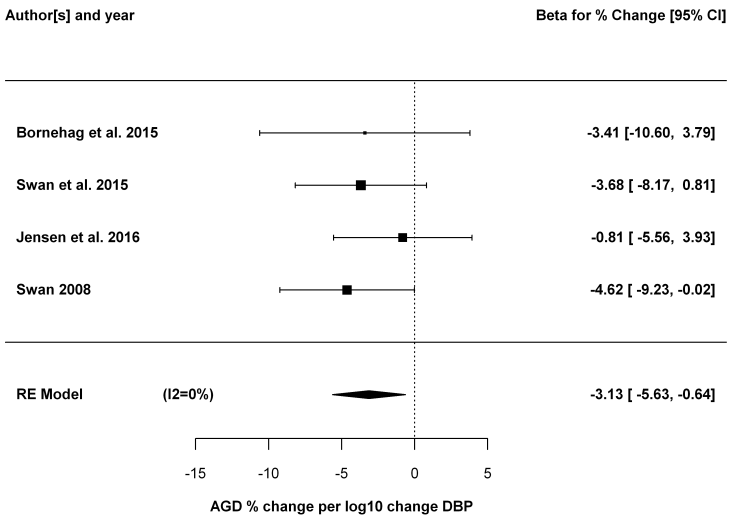


FIGURE D5-4 Meta-analysis of human studies of DBP and AGD; reported effect estimates [95% confidence interval] from individual studies and overall pooled estimate from random effects (RE) model per 10-fold increase in DBP exposure.

Sensitivity Analyses

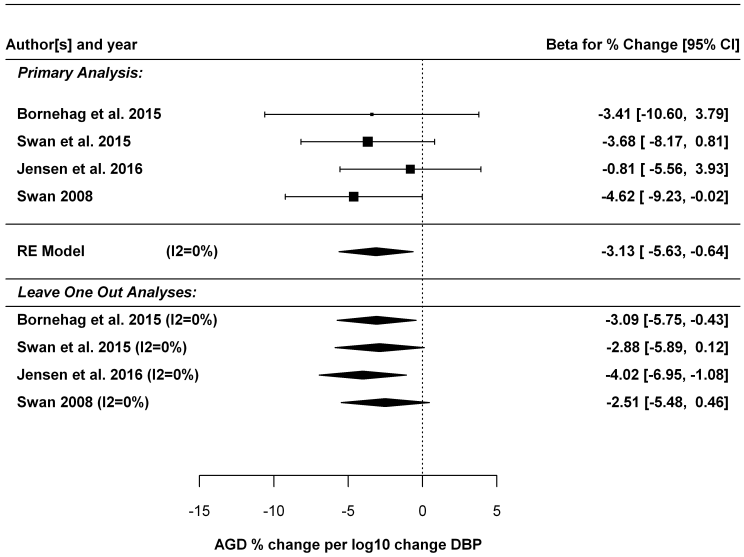


FIGURE D5-5 Sensitivity analysis of human studies of DBP and AGD performed by leaving one study out at a time.

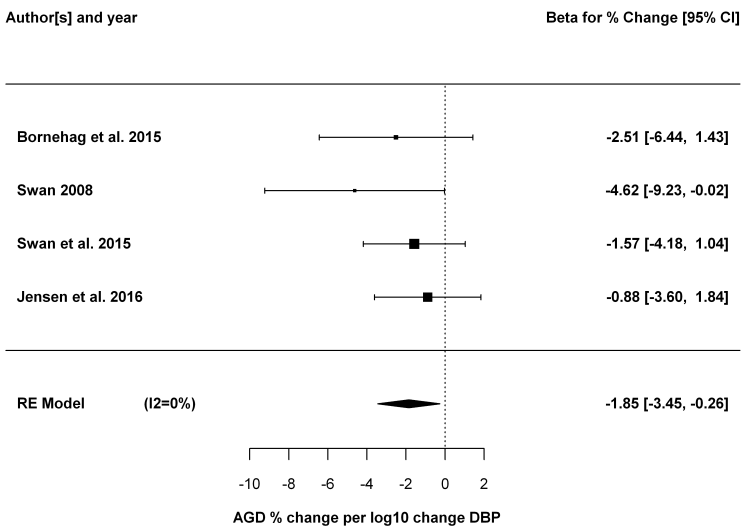


FIGURE D5-6 Sensitivity analysis of human studies of DBP and AGD performed by restricting analysis to the same outcome measure (AGD [ap]).

TABLE D5-4 Sensitivity Analyses of Human Studies of DBP and AGD

Analysis	Beta Estimate, % change	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
DBP primary analysis	-3.13	-5.63	-0.64	0.014	0.709	0
DBP w/o Bornehag et al. 2015	-3.09	-5.75	-0.43	0.023	0.502	0
DBP w/o Swan et al. 2015	-2.88	-5.89	0.12	0.060	0.522	0
DBP w/o Jensen et al. 2016	-4.02	-6.95	-1.08	0.007	0.944	0
DBP w/o Swan 2008	-2.51	-5.48	0.46	0.098	0.666	0
DBP only AGD (ap)	-1.85	-3.45	-0.26	0.023	0.566	0

DEP Meta-Analysis

Primary Analysis

In the primary analysis, four studies (see Table D5-5), with beta coefficients standardized to a percent change per $\log_{10}$ change in DEP exposure, were analyzed using a random effects model. A summary estimate of-1.94 [95% CI: -3.88, 0.001] (p = 0.0501) was found (see Figure D5-7). There was some heterogeneity, with an estimated I^2 value of 29%, though the Q statistic was not statistically significant. In the five sensitivity analyses (see Figures D5-8 and D5-9 and Table D5-6), effect sizes ranged from -1.11 to -2.54; only one of the five analyses was statistically significant. Additionally, heterogeneity with I^2>50% was observed in three of the five sensitivity analyses (though none were statistically significant). Thus, while the primary analysis suggests DEP exposure being associated with decreased AGD, the effect size is small (e.g., as compared to DEHP or DBP), the statistical significance of the result was not robust, and some heterogeneity was observed.

TABLE D5-5 Studies Included in the Meta-Analysis of DEP and AGD

Reference	Outcome	Mean AGD, mm	Exposure Metric	Estimate, %	Lower CI	Upper CI
Bornehag et al. 2015	AGD (as)	41.40	MEP (maternal urine)	1.52	-3.12	6.14
Swan et al. 2015	AGD (as)	24.73	MEP (maternal urine, trimester 1)	-1.29	-4.21	1.58
Jensen et al. 2016	AGD (as)	36.90	MEP (maternal urine)	-2.11	-5.23	1.06
Swan 2008	AGD (ap)	70.40	MEP (maternal urine)	-4.17	-7.08	-1.26

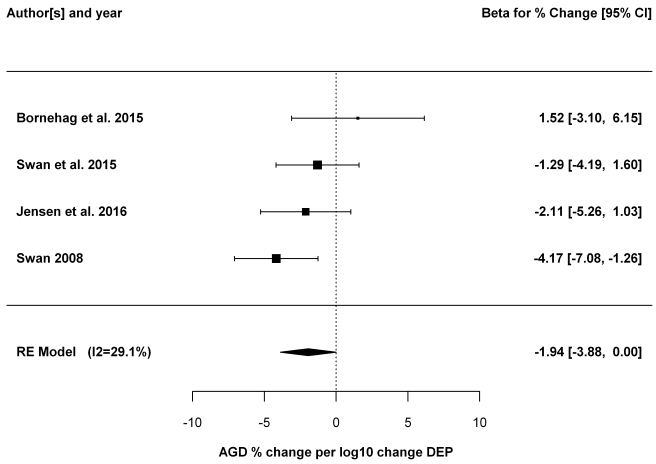


FIGURE D5-7 Meta-analysis of human studies of DEP and AGD; reported effect estimates [95% confidence interval] from individual studies and overall pooled estimate from random effects (RE) model per 10-fold increase in DEP exposure.

Sensitivity Analyses

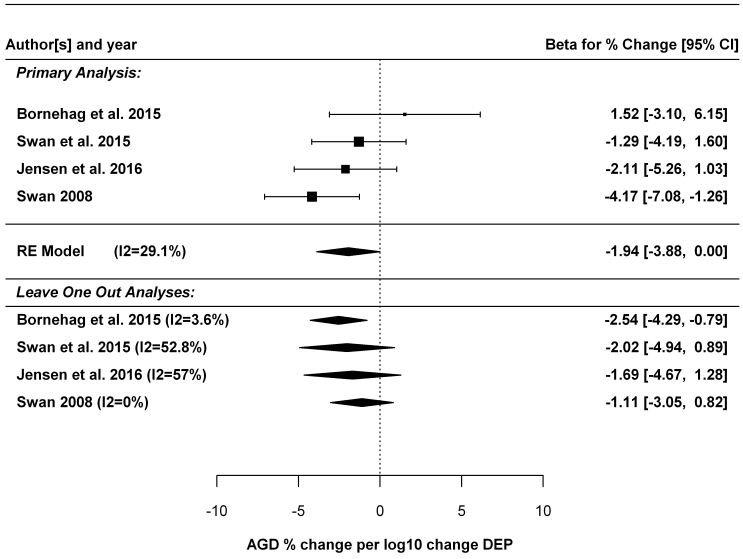


FIGURE D5-8 Sensitivity analysis of human studies of DEP and AGD performed by leaving one study out at a time.

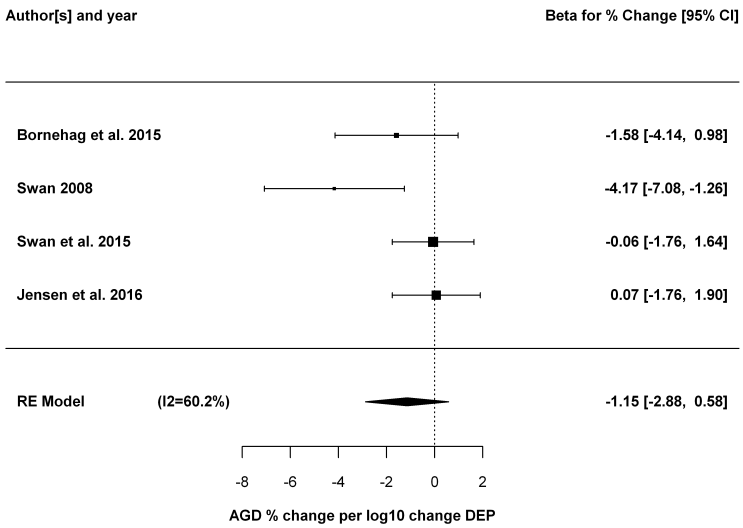


FIGURE D5-9 Sensitivity analysis of human studies of DEP and AGD performed by restricting analysis to the same outcome measure (AGD [ap]).

TABLE D5-6 Sensitivity Analyses of Human Studies of DEP and AGD

Analysis	Beta Estimate, % change	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
DEP primary analysis	-1.94	-3.88	0.001	0.050	0.204	29.1
DEP w/o Bornehag et al. 2015	-2.54	-4.29	-0.79	0.004	0.370	3.6
DEP w/o Swan et al. 2015	-2.02	-4.94	0.89	0.174	0.122	52.8
DEP w/o Jensen et al. 2016	-1.69	-4.67	1.28	0.264	0.101	57.0
DEP w/o Swan 2008	-1.11	-3.05	0.82	0.260	0.438	0.0
DEP only AGD (ap)	-1.15	-2.88	0.58	0.193	0.069	60.2

TABLE D5-7 Studies Included in the Meta-Analysis of DIBP and AGD

Reference	Outcome	Mean AGD, mm	Exposure Metric	Estimate, %	Lower CI	Upper CI
Swan et al. 2015	AGD (as)	24.73	MIBP (maternal urine, trimester 1)	-1.98	-6.83	2.91
Jensen et al. 2016	AGD (as)	36.90	MIBP (maternal urine)	-0.19	-5.61	5.18
Swan 2008	AGD (ap)	70.40	MIBP (maternal urine)	-4.20	-9.16	0.76

DIBP Meta-Analysis

In the primary analysis, three studies (see Table 5-7), with beta coefficients standardized to a percent change per $\log_{10}$ change in DIBP exposure, were analyzed using a random effects model. A summary estimate of -2.23 [95% CI: -5.15, 0.70] ($p = 0.13$) was found (see Figure D5-10). There was no significant heterogeneity, with an estimated I^2 value of 0% (Q statistic was not statistically significant). In the sensitivity analyses (see Figures D5-11 and D5-12 and Table D5-8), effect sizes ranged from -1.18 to -3.07, none of which were statistically significant. In sum, although a small effect was observed, the precision of the estimate was not sufficient to rule out chance. Thus, the available studies do not support DIBP exposure being associated with decreased AGD.

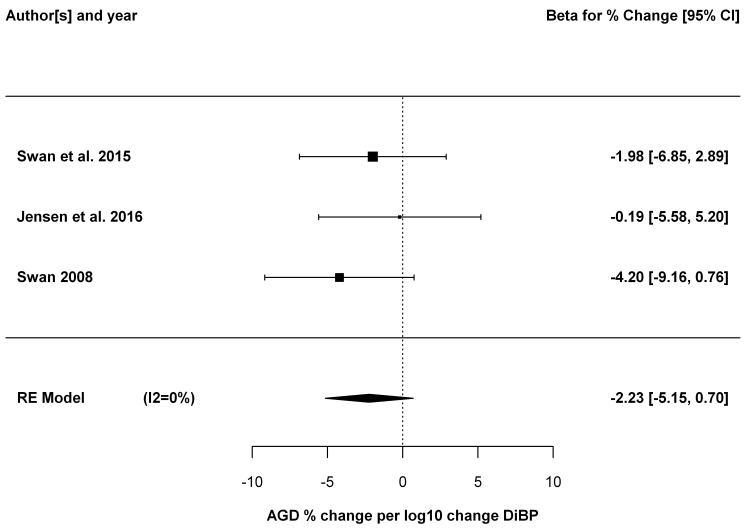


FIGURE D5-10 Meta-analysis of human studies of DIBP and AGD; reported effect estimates [95% confidence interval] from individual studies and overall pooled estimate from random effects (RE) model per 10-fold increase in DIBP exposure.

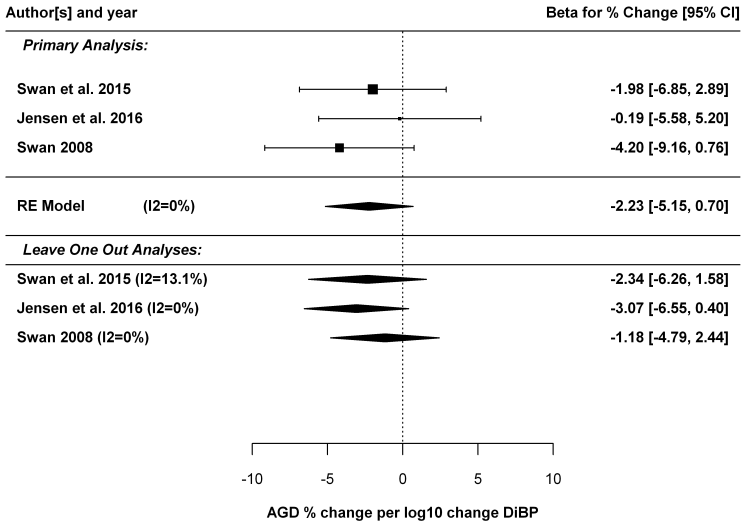


FIGURE D5-11 Sensitivity analysis of human studies of DIBP and AGD performed by leaving one study out at a time.

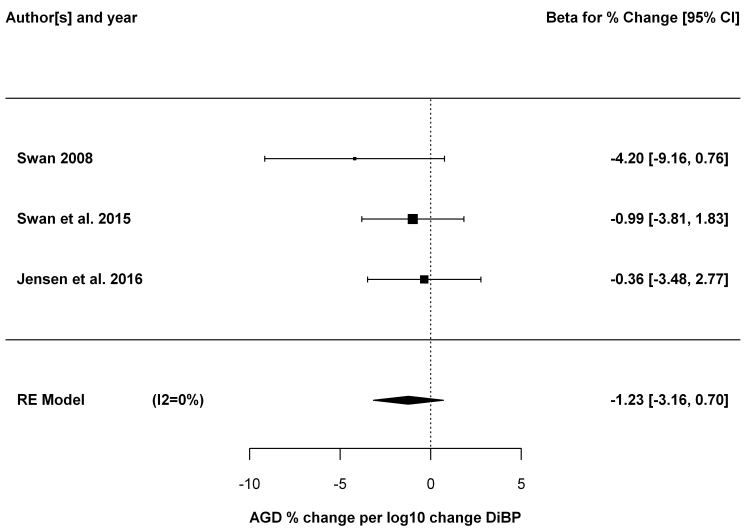


FIGURE D5-12 Sensitivity analysis of human studies of DIBP and AGD performed by restricting analysis to the same outcome measure (AGD [ap]).

TABLE D5-8 Sensitivity Analyses of Human Studies of DIBP and AGD

	Beta Estimate, % change	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
DIBP primary analysis	-2.23	-5.15	0.70	0.135	0.558	0.00
DIBP w/o Swan et al. 2015	-2.34	-6.26	1.58	0.242	0.283	13.08
DIBP w/o Jensen et al. 2016	-3.07	-6.55	0.40	0.083	0.532	0.00
DIBP w/o Swan 2008	-1.18	-4.79	2.44	0.524	0.629	0.00
DIBP only AGD (ap)	-1.23	-3.16	0.70	0.210	0.426	0.00

DINP Meta-Analysis

In the primary analysis, three studies (see Table D5-9), with beta coefficients standardized to a percent change per $\log_{10}$ change in DINP exposure, were analyzed using a random effects model. A summary estimate of-0.96 [95% CI: -4.17, 2.25] (p = 0.56) was found (see Figure D5-13). Heterogeneity was observed, with an estimated I^2 value of 58%, though the Q statistic was not statistically significant. In the sensitivity analyses (see Figures D5-14 and D5-15 and Table D5-10), effect sizes ranged from -2.42 to -0.30, none of which were statistically significant. Thus, the available studies do not support DINP exposure being associated with decreased AGD.

TABLE D5-9 Studies Included in the Meta-Analysis of DINP and AGD

Reference	Outcome	Mean AGD, mm	Exposure Metric	Estimate, %	Lower CI	Upper CI
Bornehag et al. 2015	AGD (as)	41.40	sum DINP metabolites (maternal urine)	-4.08	-8.09	-0.05
Swan et al. 2015	AGD (as)	24.73	MCOP (maternal urine, trimester 1)	1.58	-1.58	4.69
Jensen et al. 2016	AGD (as)	36.90	sum DINP metabolites (maternal urine)	-0.95	-4.69	2.74

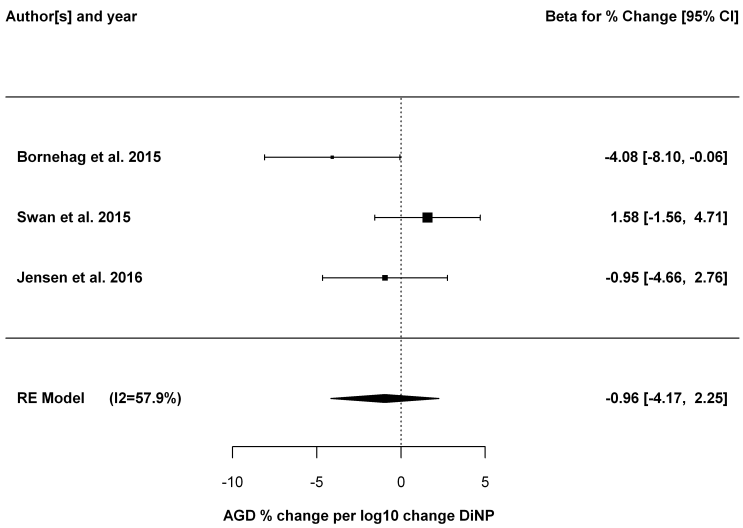


FIGURE D5-13 Meta-analysis of human studies of DINP and AGD; reported effect estimates [95% confidence interval] from individual studies and overall pooled estimate from random effects (RE) model per 10-fold increase in DINP exposure.

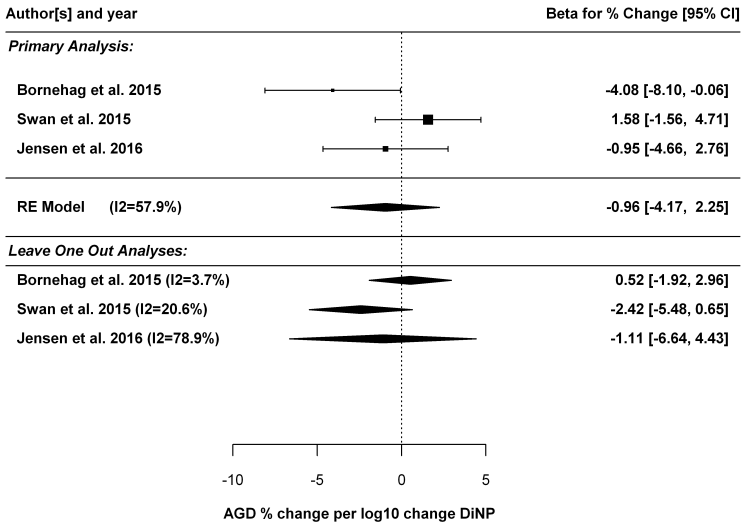


FIGURE D5-14 Sensitivity analysis of human studies of DINP and AGD performed by leaving one study out at a time.

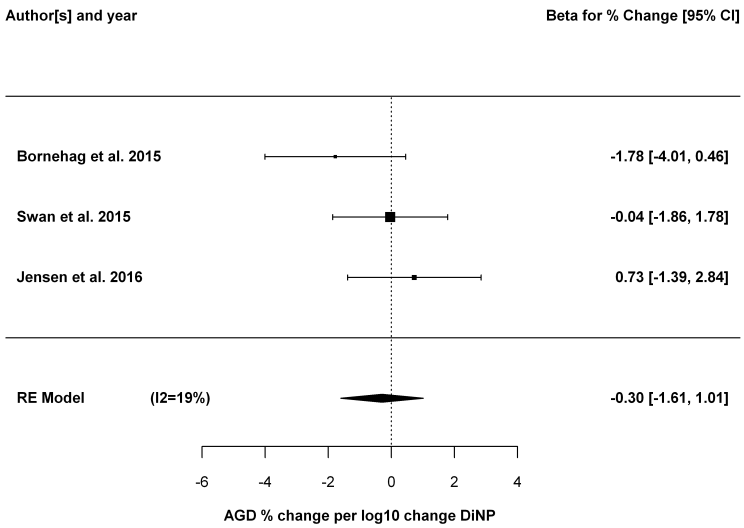


FIGURE D5-15 Sensitivity analysis of human studies of DINP and AGD performed by restricting analysis to the same outcome measure (AGD [ap]).

TABLE D5-10 Sensitivity Analyses of Human Studies of DINP and AGD

Analysis	Beta Estimate, % change	CI, Lower Bound	CI, Upper Bound	P value	P value for Heterogeneity	I^2
DINP primary analysis	-0.96	-4.17	2.25	0.559	0.092	58
DINP w/o Bornehag et al. 2015	0.52	-1.92	2.96	0.677	0.308	4
DINP w/o Swan et al. 2015	-2.42	-5.48	0.65	0.122	0.262	21
DINP w/o Jensen et al. 2016	-1.11	-6.64	4.43	0.695	0.030	79
DINP only AGD (ap)	-0.30	-1.61	1.01	0.655	0.264	19

REFERENCES

Bornehag, C.G., F. Carlstedt, B.A. Jonsson, 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. Environ. Health Perspect. 123(1):101-107.

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. J. Dev. Orig. Health Dis. 4(4):300-306.

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, and A.M. Andersson. 2016. Prenatal exposure to phthalates and anogenital distance in male infants from a low-exposed Danish cohort (2010-2012). Environ. Health Perspect. 124(7):1107-1113.

Martino-Andrade, A.J., F. Liu, S. Sathyanarayana, E.S. Barrett, J.B. Redmon, R.H. Nguyen, H. Levine, and S.H. Swan. 2016. Timing of prenatal phthalate exposure in relation to genital endpoints in male newborns. Andrology 4(4):585-593.

Swan, S.H. 2008. Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. Environ. Res. 108(2):177-184.

Swan, S.H., S. Sathyanarayana, E.S. Barrett, S. Janssen, F. Liu, R.H. Nguyen, and J.B. Redmon. 2015. First trimester phthalate exposure and anogenital distance in newborns. Hum. Reprod. 30(4):963-972.

Appendix E

Supporting Materials for the PBDE (Animal) Systematic Review

SECTION E-1

PBDE (ANIMAL) SYSTEMATIC REVIEW PROTOCOL

August 3, 2016
(Modified on September 15, 2016—See Section E-1f)

BACKGROUND AND INTRODUCTION

Polybrominated diphenyl ethers (PBDEs) are synthetic brominated flame retardants that are ubiquitous environmental contaminants that have been measured in animals and in humans. Because the developing organism has been shown to be particularly vulnerable to endocrine-disrupting chemicals, such as PBDEs, the committee decided to focus on studies of developmental exposure. PBDEs have been linked to effects on neurodevelopment after developmental exposure in animal studies.

OBJECTIVE AND SPECIFIC AIMS

Review Question

The overall objective of this systematic review is to answer the question is developmental exposure to PBDEs in nonhuman mammals associated with alterations in learning, memory, attention, or response inhibition?

The specific aims of the review are to:

- Identify literature reporting the effects of developmental PBDE exposure on learning, memory, attention, or response inhibition in nonhuman mammals.
- Extract data on learning, memory, attention, or response inhibition from relevant studies.
- Assess the internal validity (risk of bias) of individual studies.
- Summarize the extent of evidence available.
- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration, such as study-design heterogeneity.
- Rate the confidence in the body of evidence for studies in nonhuman mammals according to one of five statements: (1) high; (2) moderate; (3) low; (4) very low/no evidence available; or (5) evidence of lack of neurotoxicity.

PECO Statement

A PECO (Population, Exposure, Comparator, and Outcome) statement was developed by the review team as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the review question for the systematic review.

Population: Nonhuman mammals

Exposure:

- PBDE refers to any single PBDE congener or combination of grouped congeners.
- Any developmental exposure to PBDEs, with no restrictions based on route of exposure or administered dose or concentration. To be considered "developmental," the exposure occurred during any of the following periods: prior to conception in one or both parents, prenatal in the pregnant female (exposure to offspring in utero), or postnatal until sexual maturation.

Comparator: Nonhuman mammals exposed during development to different doses of PBDEs or vehicle-only treatment.

Outcomes: Measures of learning, memory, attention, or response inhibition. Examples of tests include Morris water maze, radial arm maze, and operant tests of cognition.

METHODS

Problem Formulation and Protocol Development

The review question and specific aims were developed and refined through a series of problem formulation steps. The committee considered review articles on endocrine disruptors in surveying the types of chemicals that might make good case examples and held a workshop to explore potential case examples. The committee sought an example of a chemical for which both the human and the animal evidence on effects appears to be associated with different exposure levels of that chemical and due to perturbation of the estrogen or androgen hormone system. PBDEs appeared to fit this case criterion.

The protocol will be peer reviewed by subject-matter and systematic-review experts in accordance with standard report-review practices of the National Academies of Sciences, Engineering, and Medicine. The protocols will be revised in response to peer review comments and will subsequently be published as appendices to the committee's final report. The identity of the peer reviewers will remain anonymous to the committee until the publication of the final report, when their names and affiliations are disclosed in the Preface.

Committee and Staff

There are 11 committee members, supported by two staff members of the National Academies. The committee members were appointed in accordance with the standard policies and practices of the National Academies on the basis of their expertise in general toxicology, reproductive toxicology, developmental toxicology, endocrinology, neurotoxicology, epidemiology, risk assessment, biostatistics, and systematic-review methods. The membership of the committee and the staff was determined before the topic of the systematic review was selected. It was known, however, that each case study would be on an endocrine-disrupting chemical, so committee members who have relevant expertise were specifically recruited and appointed.

Review Team

The review team for this case study will be a subgroup of the committee (BH, SSc), two National Academies staff members (EM, SM), and an information specialist (JB). If a member of the review team was a coauthor of a study under review, that member will recuse himself or herself from the evaluation of the quality of that study.

The review team will be responsible for performing all aspects of the review, including conducting the literature searches; applying inclusion/exclusion criteria to screen studies; extracting data; assessing risk of bias for included studies; and analyzing and synthesizing data. The roles and responsibilities of the team members will be documented throughout the protocol. Throughout the course of its work, the review team will also engage other members of the committee to provide consultation as needed. The involvement of those individuals will be documented and acknowledged.

Biographical information on the review team is presented in Section E-1a.

Search Methods

Search for Existing Systematic Reviews

The review team will consider using existing systematic reviews to address or help to address its research question. English-language systematic reviews conducted within the last 3 years will be sought. The review team will incorporate prior reviews, update prior reviews, and/or use the reviews as part of its searching, depending on determination of their relevancy and quality (Whitlock et al. 2008). Current guidance on using existing systematic reviews will be used (Robinson et al. 2014, 2015, 2016).

Search

Recent, relevant high-quality systematic reviews addressing the research question about PBDEs and neurodevelopment will be searched. PubMed will be searched by adding the qualifier "systematic review"[ti] OR "meta-analysis"[pt] OR "meta-analysis"[ti] OR ("systematic"[ti] AND "review"[ti]) OR (systematic review [tiab] AND review [pt]) OR "meta synthesis"[ti] OR "meta synthesis"[ti] OR "integrative review"[tw] OR "integrative research review"[tw] OR "cochrane database syst rev"[ta] OR "evidence synthesis"[tiab] to the preliminary search strategy (see Section E-1b). Language and date restrictions will be applied (English language; published 2013 to present). The systematic review protocol registries PROSPERO (CRD) and CAMARADES will also be searched using key terms from the preliminary PubMed strategy.

Study Selection

Two team members (SM, EM) will independently screen search results, applying the following exclusion criteria:

- Not a systematic review.[1] The minimum criteria for a study to be considered a systematic review are
  - conduct of an explicit and adequate literature search,
  - application of predefined eligibility criteria,
  - consideration of the quality of included studies or risk of bias assessment, and
  - synthesis (or attempt at synthesis) of the findings, either qualitatively or quantitatively.

[1]A systematic review "is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

- Not in English.
- Search date prior to 2013.
- Does not match the research question or PECO elements.

For PubMed results, screening will be conducted first using abstracts and then at the full-text level. Results from PROSPERO and CAMARADES will be conducted at one level, using the information in the registry. Disagreements regarding eligibility will be resolved through discussion or, where necessary, by a third team member.

Assessment for Quality

Two investigators (KR, AR) will independently assess the risk of bias of eligible systematic reviews using ROBIS (Whiting et al. 2016). Disagreements in rating will be resolved through discussion or, where necessary, through consultation with a third team member. Systematic reviews rated as low quality will be excluded from further consideration at this stage.

Use of Existing Reviews

Eligible systematic reviews of high quality will be reviewed, considering date of search, match with the PECO statement, as well as availability of data from the primary studies, how risk of bias was conducted, and other factors. Current reviews considered a good match will be used to address the research question. Reviews that are a good match but with search dates more than a year ago will be updated. If no relevant systematic reviews are found, an independent systematic review will be performed.

Literature Search for Independent Systematic Review

The review team will collaborate with an information specialist (JB) who has training, expertise, and familiarity with developing and performing systematic review literature searches. A variety of methods will be used to identify relevant data (see below). Literature searches will not be limited by publication date.

Online Databases

Electronic searches of the following three online databases will be performed using the search terms outlined in Section E-1b: PubMed, Embase, and Toxline. The search strategy and search terms will be developed by the information specialist (JB), who will implement the search for relevant studies.

Other Resources

Hand searching the reference lists of all the included studies after full-text review will be conducted using the same study selection process as was used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

Study Selection

All search results will be imported or manually entered into EndNote (Version x7) reference management software. EndNote will be used to eliminate any duplicate citations before evaluating the eligibility of the citations.

Screening Process

References retrieved from the literature search will be screened for relevance and eligibility against the evidence selection criteria using DistillerSR (Evidence Partners; https://www.evidencepartners.com). Screeners from the review team will be trained with an initial pilot phase on 25 studies undertaken to improve clarity of the evidence selection criteria and to improve accuracy and consistency among screeners. Screening forms are presented in Section E-1c.

Title and Abstract Screening

Each citation will be independently screened by two reviewers (SM, EM) to determine whether it meets the selection criteria for inclusion that reflect the PECO statement with some additional considerations as listed below. Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

The title/abstract screening form will be used to screen and EXCLUDE references if at least one of the following criteria is met:

1. No original data (e.g., review article, commentary, editorial)
2. Study does not include nonhuman mammals
3. Study does not report PBDE exposure
4. No relevant outcomes
5. Incomplete information (e.g., conference abstract, meeting poster)
6. Not in English and unable to determine eligibility
7. Other (explanation required)

The following types of records will be INCLUDED at the title/abstract level: any English-language study of nonhuman mammals exposed to PBDEs.

Only English-language publications will be included, because of time and resource constraints. There appears to be no indication that foreign-language publications would make a contribution that is distinct from what is found in the English-language literature.

Updated details to instructions and interpretations for title and abstract screening will be added to Section E-1f to document the process of the review team during the screening process.

Full-Text Screening

Citations included at the title/abstract screening level will be subject to a full-text review by the same two reviewers involved in title and abstract screening (SM, EM). Each reference will be screened in duplicate and independently. Disagreements regarding citation eligibility will be resolved via consensus and, where necessary, by consulting a committee member.

Citations will be EXCLUDED at the full-text level if at least one of the following criteria is met:

1. No original data (e.g., review article, commentary, editorial)
2. Study does not include nonhuman mammals
3. Study does not report experimental PBDE exposure
4. Study does not quantify exposure to PBDE
5. Study does not include developmental exposure (prior to conception in one or both parents, prenatal in the pregnant female [exposure to offspring in utero], or postnatal until sexual maturation)
6. Study does not assess or report quantitative measures of learning, memory, attention, or response inhibition
7. No comparator group (animals exposed to different doses of PBDEs or vehicle-only treatment)
8. Not in English

9. Other reason (explanation required)

The reason for exclusion at the full-text-review stage will be annotated and reported in a study selection flow diagram in the final report (following PRISMA [Moher et al. 2009]). The reasons for exclusion will be documented from the list (1-9) above.

Citations will be INCLUDED if they meet the PECO statement criteria:

- Study includes nonhuman mammals
- Study includes developmental exposure
- Study includes comparison with animals exposed to different doses of PBDEs or vehicle-only treatment
- Study measures (1) learning, (2) memory, (3) attention, or (4) response inhibition.

Updated details to instructions and interpretations for full-text screening will be added to Section E-1f to document the process of the review team during the screening process.

Data Extraction

Data will be collected and recorded (i.e., extracted) from included studies by one member of the review team and checked by a second member for completeness and accuracy. Any discrepancies in data extraction will be resolved through discussion. The extracted data will be used to summarize study designs and findings and/or to conduct statistical analyses. Section E-1d presents the data extraction elements that will be used.

The review team will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The study extraction files will note whether an attempt was made to contact study authors by email for missing data considered important for evaluating key study findings (and whether or not a response was received).

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 review will include all publications on the study, select one publication to use as the primary publication, and consider the others as secondary publications with the annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up or, for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. The review will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, the review team will include a single instance of the data (and avoid more than one; that is, duplicate instances of the data).

Data extraction will be completed using the Health Assessment Workspace Collaborative (HAWC) software, an open source and freely available Web-based interface application, for visualization and warehousing.[2]

Risk of Bias (Quality) Assessment of Individual Studies

Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and

[2]HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

even the direction of the apparent effect. Internal validity or risk of bias will be assessed for individual studies using a tool developed by the National Toxicology Program's Office of Health Assessment and Translation (OHAT) that outlines an approach to evaluating risk of bias for experimental animal studies. The risk of bias domains and questions for experimental animal studies are based on established guidance for experimental human studies (randomized clinical trials) (Viswanathan et al. 2012, 2013; Sterne et al. 2014; Higgins and Green 2011) and recent tools for animal studies (Hooijmans et al. 2014; Koustas et al. 2014). The risk of bias tool includes a common set of questions (Section E-1e) that are answered based on the specific details of individual studies to develop risk of bias ratings (using the four options: definitely low risk of bias; probably low risk of bias; probably high risk of bias; or definitely high risk of bias). Information or study procedures that were not reported are assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Study design determines the subset of questions that should be used to assess risk of bias for an individual study (see Table E1-1).

Studies are independently assessed by two assessors (BH, SSc) who answer all applicable risk of bias questions with one of four options (see Table E1-2) following prespecified criteria detailed in Section E-1e. The criteria describe aspects of study design, conduct, and reporting required to reach risk of bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific type of human study designs. Risk of bias will be assessed at the outcome level because study-design or method specifics may increase the risk of bias for some outcomes and not others within the same study. Information or study procedures that were not reported are assumed not to have been conducted, resulting in an assessment of "probably high" risk of bias. Authors will be queried by email to obtain missing information, and responses received will be used to update risk of bias ratings.

Assessors will be trained using the criteria in an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk of bias assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk of bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk of bias criteria.

After assessors have independently made risk of bias determinations for a study across all risk of bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered and resolved with the review team. The final risk of bias rating for each question will be recorded along with a statement of the basis for that rating.

Data Analysis and Evidence Synthesis

The review team will qualitatively synthesize the body of evidence for each outcome and, where appropriate, a meta-analysis will be performed. If a meta-analysis is performed, summaries of main characteristics for each included study will be compiled and reviewed by two team members to determine comparability between studies, to identify data transformations necessary to ensure comparability, and to determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

- Experimental design (e.g., acute, chronic, multigenerational)
- Animal model used (e.g., species, strain, sex, genetic background)
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status)

TABLE E1-1 OHAT Risk of Bias Tool

Risk-of-Bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Nonhuman animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.

**Human Controlled Trials are studies in humans with controlled exposure (e.g., randomized controlled trials, nonrandomized experimental studies).

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

SOURCE: NTP (2015, p. 37).

TABLE E1-2 Answers to the Risk of Bias Questions

++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices.
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias.
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices.

SOURCE: NTP (2015, p. 36).

- Developmental stage of animals at treatment and outcome assessment
- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

The review team expects to require input from subject-matter experts to help assess the heterogeneity of the studies. Subgroup analyses to examine the extent to which risk of bias contributes to heterogeneity will be performed. If a meta-analysis is considered appropriate, the review team will stratify by species and further consider separate meta-analyses by species. Situations where it may not be appropriate to include a study are when data on exposure or outcome are too different to be combined or other circumstances that may indicate that averaging study results would not produce meaningful results. When considering outcome measures for conducting meta-analyses, benchmark dose (BMD) estimates (and their associated confidence intervals) with a benchmark response (BMR) set to a common percent of control (for continuous outcomes) or extra risk (for dichotomous outcomes) are preferred. A secondary alternative, when there are more than two groups, is the conduct of BMD modeling and the use of the derived BMD estimates. Meta-analyses are not possible with lowest-observed-adverse-effect levels or no-observed-adverse-effect levels, since no confidence interval can be derived for these measures.

If a meta-analysis is conducted, a random effects model will be used for the analysis. Heterogeneity will be assessed using the I-squared statistic. Interpretation of I-squared will be based on the Cochrane Handbook: 0% to 40% (might not be important); 30% to 60% (may represent moderate heterogeneity); 50% to 90% (may represent substantial heterogeneity); and 75% to 100% (considerable heterogeneity). Additionally, as described in the Cochrane Handbook, for the last three categories, the importance of the I-squared will be interpreted considering not only the magnitude of effects but also the strength of the evidence (90% two-tailed confidence interval).

The review team will also perform sensitivity analyses on the exclusion of individual studies in succession.

If sufficient studies are available, subgroup analyses will be performed based on the following characteristics described above: experimental design, animal model used (e.g., species and/or strain), age of animals, and developmental stage of animals at treatment and outcome assessment.

In the event that these proposed methods for data analysis are altered to tailor to the evidence base from included studies, the protocol will be amended accordingly, and the reasons for change will be justified in the documentation.

Confidence Rating: Assessment of the Body of Evidence

The quality of evidence for each outcome will be evaluated using the GRADE system for rating the confidence in the body of evidence (Guyatt et al. 2011; Rooney et al. 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in NTP (2016, see Step 5). In brief, available studies on a particular outcome are initially grouped by key study-design features, and each grouping of studies is given an initial confidence rating by those features.

The initial rating is downgraded for factors that decrease confidence in the results, including

- risk of bias
- unexplained inconsistency
- indirectness or lack of applicability
- imprecision
- publication bias

The initial rating is upgraded for factors that increase confidence in the results, including

- large magnitude of effect
- dose response
- consistency across study designs/populations/animal models or species
- consideration of residual confounding
- other factors that increase confidence in the association or effect (e.g., particularly rare outcomes)

The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt et al. 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney et al. 2014).

Confidence ratings are independently assessed by members of the review team, and discrepancies will be resolved by consensus and consultation with technical advisors as needed. Confidence ratings will be summarized in evidence profile tables.

REFERENCES

Guyatt, G.H., A.D. Oxman, R. Kunz, J. Brozek, P. Alonso-Coello, D. Rind, P.J. Devereaux, V.M. Montori, B. Freyschuss, G. Vist, R. Jaeschke, J.W. Williams, Jr., M.H. Murad, D. Sinclair, Y. Falck-Ytter, J. Meerpohl, C. Whittington, K. Thorlund, J. Andrews, and H.J. Schunemann. 2011. GRADE guidelines 6. Rating the quality of evidence—imprecision. J. Clin. Epidemiol. 64(12):1283-1293.

Higgins, J., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available: http://handbook.cocharne.org [accessed May 6, 2016].

Hooijmans, C.R., M.M. Rovers, R.B. de Vries, M. Leenars, M. Ritskes-Hoitinga, and M.W. Langendam. 2014. SYRCLE's risk of bias tool for animal studies. BMC Med. Res. Method. 14:43.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Koustas, E., J. Lam, P. Sutton, P.I. Johnson, D.S. Atchley, S. Sen, K.A. Robinson, D.A. Axelrad, and T.J. Woodruff. 2104. The Navigation Guide—evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122(10):1015-1027.

Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. J. Clin. Epidemiol. 62(10):1006-1012.

NTP (National Toxicology Program). 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation, Division, National Toxicology Program, National Institute of Environmental Health Sciences. January 9, 2015 [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed September 21, 2015].

Robinson, K.A., E.P. Whitlock, M.E. O'Neil, J.K. Anderson, L. Hartling, D.M. Dryden, M. Butler, S.J. Newberry, M. McPheeters, N.D. Berkman, J.S. Lin, and S. Chang S. 2014. Integration of Existing Systematic Reviews. Research white paper. AHRQ Publication No. 14-EHC016-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK216379/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2015. Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 15-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality [online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279904/ [accessed May 9, 2016].

Robinson, K.A., R. Chou, N.D. Berkman, S.J. Newberry, R. Fu, L. Hartling, D. Dryden, M. Butler, M. Foisy, J. Anderson, M. Motu'apuaka, R. Relevo, J.M. Guise, and S. Chang. 2016. Twelve recommendations for integrating existing systematic reviews into new reviews: EPC guidance. J. Clin. Epidemiol. 70:38-44.

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. Environ. Health Perspect. 122(7):711-718.

Sterne, J.A.C., J.P.T. Higgins, and B.C. Reeves. 2014. ACROBAT-NRSI: A Cochrane risk of Bias Assessment Tool for Non-randomized Studies of Interventions. Version 1.0.0. September 24, 2014 [online]. Available: www.riskofbias.info [accessed May 6, 2016].

Viswanathan, M., M. Ansari, N.D. Berkman, S. Chang, L. Hartling, L.M. McPheeters, P.L. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze, and J.R. Treadwell. 2012. Assessing the Risk of Bias of Individual Studies when Comparing Medical Interventions. AHRQ Publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available www.effectivehealthcare.ahrq.gov/ [accessed May 6, 2016].

Viswanathan, M., N.D. Berkman, D.M. Dryden, and L. Hartling. 2013. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank. Methods Research Report. AHRQ Publication No. 13-EHC106-EF. Rockville, MD: Agency for Healthcare Research and Quantitative Methods Guide for Comparative Effectiveness Reviews [online]. Available www.effectivehealthcare.ahrq.gov/reports/final.cfm [accessed May 6, 2016].

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

Whitlock, E.P., J.S. Lin, R. Chou, P. Shekelle, and K.A. Robinson. 2008. Using existing systematic reviews in complex systematic reviews. Ann. Intern. Med. 148(10):776-782.

SECTION E-1a

REVIEW TEAM BIOGRAPHICAL INFORMATION

Jaime F. Blanck is a clinical informationist at the Welch Medical Library at Johns Hopkins University. She creates and implements systematic review search strategies across multiple databases and provides comprehensive reference, research, and information services to multiple departments within the School of Medicine. She received an MLIS from the University of Pittsburgh and an MPA from the University of Baltimore.

Barbara F. Hales is a James McGill Professor in the Department of Pharmacology and Therapeutics at McGill University. Her research interests are in the mechanisms of action of drugs as teratogens. She studies developmental toxicity using a combination of in vivo, in vitro, and molecular approaches with the goal of elucidating how the embryo responds to insult after direct or maternal exposure and the consequences to progeny of paternal drug exposure. Dr. Hales is a past president of the Teratology Society, and is currently co-chair of the Chemicals Management Plan Science Committee of the Government of Canada. She received an MSc in pharmacognosy from the Philadelphia College of Pharmacy and Science and a PhD in pharmacology and therapeutics from McGill University.

Ellen Mantus is a scholar and director of risk assessment on the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine with more than 20 years of experience in the fields of toxicology and risk assessment. She has served as the study director on numerous projects, including ones that have assessed the health implications of various chemical exposures; developed strategies for applying modern scientific approaches in toxicology and risk assessment; provided guidance to federal agencies on risk-based decision making; and evaluated barriers to deployment of electric vehicles and associated charging infrastructure. Before joining the National Academies, Dr. Mantus was a project manager with ICF Consulting where she served as a primary reviewer for numerous toxicological studies and provided risk assessment and regulatory support on a wide array of projects. Dr. Mantus received a PhD in chemistry from Cornell University.

Susan Martel is a senior program officer in the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine. She has 20 years of experience in supporting toxicology and risk assessment projects for the US Environmental Protection Agency, the US Department of Defense, and the National Aeronautics and Space Administration. Recent projects include working with committees evaluating the toxicological effect of arsenic, developing exposure guidelines for use on spacecraft, and assessing pesticide risks-assessment practices. Before joining the National Academies, she was the administrator of the Registry for Toxicology Pathology for Animals at the American Registry of Pathology. She received a BA in biology from Skidmore College.

Susan L. Schantz is a professor of toxicology in the Department of Comparative Biosciences, College of Veterinary Medicine, at the University of Illinois at Urbana-Champaign. She is also director of a National Institute of Environmental Health Sciences (NIEHS) T32 training program in endocrine, developmental, and reproductive toxicology and director of a Children's Environmental Health Research Center jointly funded by the NIEHS and the EPA. In addition, she is currently the interim director of the Neuroscience Program. Dr. Schantz's research interests involve understanding the neurobehavioral effects of chemical exposures during development and aging. She conducts research in both laboratory-based animal studies and parallel epidemiologic studies. She has served as president of the Neurotoxicology Specialty Section of the Society of Toxicology and president of the Neurobehavioral Teratology Society. Dr. Schantz was also a member of the NRC's Committee to Assess the Health Implications of Perchlorate Ingestion. She received a PhD in environmental toxicology from the University of Wisconsin–Madison.

SECTION E-1b

LITERATURE SEARCH STRATEGY

The review team will employ a multi-method process to identify all potentially relevant studies as detailed below.

Electronic Searches

PubMed

A search string employing medical subject heading (MeSH) terms and keyword synonyms will be developed. The PubMed search strategy will be considered the primary search strategy and will provide the basis of the other electronic search strategies. To assist in compiling these terms, the review team will consult an existing systematic review protocol studying PBDEs in humans (J. Lam et al. Applying the navigation guide systematic review methodology. Case study #5: association between developmental exposures to PBDEs and human neurodevelopment. PROSPERO 2015:CRD42015019753 Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42015019753). This protocol was selected because it examines the substances of interests and timing of exposure in a parallel human population. The search strategies will address each of the following concepts:

- *Flame retardants (PBDEs)*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the *Flame retardants (PBDEs)* concept. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. CAS registry numbers for each PBDE substance will also be included in the list of search terms. All MeSH terms, Supplementary Concept terms, keyword synonyms, and CAS registry numbers will be searched together as one concept using the Boolean operator "OR."
- *Exposure*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the *exposure* concept. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. All MeSH terms and keyword synonyms will be searched together as one concept using the Boolean operator "OR."
- *Animal studies*—The review team will adapt the search filter published in Hooijmans CR, Tillema A, Leenaars M, Ritsk*es-Hoitinga M. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Laboratory Animals. 2010;44(3):170-175* to eliminate nonmammalian animals*. doi:10.1258/la.2010.009117.*
- *Outcomes*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to measures of learning, memory, attention, and cognition. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. All MeSH terms and keyword synonyms will be searched together as one concept using the Boolean operator "OR."

Each of the above concepts will be searched together using the Boolean operator "AND." There will not be limitations on date of publication, language, or publication type. All citation records will be exported to EndNote. Additional citations identified through the search processes identified below will also be exported to the project EndNote library. Duplicates will be removed from the citation library using the "Find Duplicates" tool in EndNote as well as a manual review of citations by the project librarian to identify any duplicates not found during the automated process. The number of citations found in

each database will be recorded as well as the number of duplicates and final tally of unique citations. The final library of citations will be uploaded to the Health Assessment Workspace Collaboration Web-based tool (www.hawcproject.org) for systematic reviews where they will be reviewed by the team.

Embase

The controlled vocabulary database Emtree is used by Embase. For each MeSH term identified through process above, Emtree will be searched for the appropriate corresponding term. Additional keywords will identified using the list of synonyms from each Emtree record and added to the keywords from the MeSH records. The review team will substitute the animal study search filter used in the PubMed search with the comparable Embase filter published in *De Vries RBM, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. A search filter for increasing the retrieval of animal studies in Embase.Laboratory Animals. 2011;45(4):268-270. doi:10.1258/la.2011.011056.* This version of the animal filter will also be adapted to remove all nonmammalian animals.

Toxline

The review team will develop the Toxline search strategy by removing any database specific formatting from the PubMed search strategy to create a keyword-only search (Toxline does not employ a controlled vocabulary).

Search Strategies

PubMed

("Flame Retardants"[Mesh] OR"Flame Retardants" [Pharmacological Action] OR "Halogenated Diphenyl Ethers"[Mesh] OR ("Phenyl Ethers"[Mesh:NoExp] AND ("1974/01/01"[PDAT] : "2008/12/31"[PDAT])) OR "pentabromodiphenyl ether" [Supplementary Concept] OR "2,2',3,3',4,4',6,6'-octabromodiphenyl ether" [Supplementary Concept] OR "decabromobiphenyl ether" [Supplementary Concept] OR "tribromodiphenyl ether 28"[Supplementary Concept] OR "2,2',4,4'-tetrabromodiphenyl ether" [Supplementary Concept] OR "2,2',4,5'-tetrabromodiphenyl ether" [Supplementary Concept] OR "hexabromodiphenyl ether 154" [Supplementary Concept] OR "2,2',4,4',5,6'hexabromodiphenyl ether" [Supplementary Concept] OR "2,2',3,4,4',5',6heptabromodiphenyl ether" [Supplementary Concept] OR "2,2',3,3',4,5,5',6,6'-nonabromodiphenyl ether" [Supplementary Concept] OR "2,2',3,3',4,4',5,6,6'-nonabromodiphenyl ether" [Supplementary Concept] OR "2,2',3,3',4,4',5,5',6-nonabromodiphenyl ether" [Supplementary Concept] OR "2,2',4,4',5,5'-hexabrominated diphenyl ether" [Supplementary Concept] OR "hexabrominated diphenyl ether 153" [Supplementary Concept] OR "pentabrominated diphenyl ether 100" [Supplementary Concept] OR "5-OH-BDE-47" [Supplementary Concept] OR "6-OH-BDE-47" [Supplementary Concept] OR flame retard*[tw] OR fire retard*[tw] OR fireproofing agent*[tw] OR "FireMaster"[tw] OR "Bromkal"[tw] OR diphenyl ether deriv*[tw] OR halogenated diphenyl*[tw] OR brominated diphenyl*[tw] OR PBDE*[tw] OR polybrominated diphenyl*[tw] OR polybromodiphenyl*[tw] OR PBDP*[tw] OR BDE*[tw] OR pentabromodiphenyl*[tw] OR cpentaBDE*[tw] OR PentaBDE*[tw] OR "PeBDE"[tw] OR "DE 71"[tw] OR "DE71"[tw] OR "pentabrominated diphenyl"[tw] OR "pentabrominated diphenyls"[tw] OR "PBDPO"[tw] OR "Planelon PB 501"[tw] OR pentabromo deriv*[tw] OR pentabromophenyl*[tw] OR octabromodiphenyl*[tw] OR c-octaBDE*[tw] OR OctaBDE*[tw] OR "OcBDE"[tw] OR "Octabrom"[tw] OR octabromo deriv*[tw] OR "OBDE"[tw] OR "OBDPO"[tw] OR "octabrominated diphenyl"[tw] OR "octabrominated diphenyls"[tw] OR decabromodiphenyl*[tw] OR cdecaBDE*[tw] OR DecaBDE*[tw] OR "DeBDE"[tw] OR "DBDPO"[tw] OR "decabrominated diphenyl"[tw] OR "decabrominated diphenyls"[tw] OR decabromo deriv*[tw] OR "Decabrom"[tw] OR "Berkflam B 10E"[tw] OR "FR 300BA"[tw] OR "FR 300 BA"[tw] OR tribromodiphenyl*[tw] OR "tribrominated diphenyl"[tw] OR "tribrominated diphenyls"[tw] OR "TrBDE"[tw] OR tribromo deriv*[tw] OR tetrabromodiphenyl*[tw] OR TetraBDE*[tw] OR "TeBDE"[tw]

OR "TBDE"[tw] OR "BPDE"[tw] OR tetrabromo deriv*[tw] OR "TBDP"[tw] OR "tetrabrominated diphenyl"[tw] OR "tetrabrominated diphenyls"[tw] OR hexabromodiphenyl*[tw] OR HexaBDE*[tw] OR "HxBDE"[tw] OR "hexabrominated diphenyl"[tw] OR "hexabrominated diphenyls"[tw] OR hexabromo deriv*[tw] OR heptabromodiphenyl*[tw] OR HeptaBDE*[tw] OR "HeBDE"[tw] OR "heptabrominated diphenyl"[tw] OR "heptabrominated diphenyls"[tw] OR heptabromo deriv*[tw] OR nonabromodiphenyl*[tw] OR NonaBDE*[tw] OR "NoBDE"[tw] OR "nonabrominated diphenyl"[tw] OR "nonabrominated diphenyls"[tw] OR nonabromo deriv*[tw] OR "7025-06-1" OR "6876-00-2" OR "101-55-3" OR "51452-870" OR "446254-14-4" OR "147217-72-9" OR "171977-449" OR "147217-71-8" OR "33513-66-3" OR "51930-04-2" OR "6903-63-5" OR "189084-59-1" OR "83694-71-7" OR "46438-88-4" OR "2050-47-7" OR "147217-74-1" OR "147217-75-2" OR "407606-55-7" OR "147217-73-0" OR "147217-76-3" OR "337513-67-4" OR "446254-15-5" OR "446254-16-6" OR "147217-77-4" OR "337513-75-4" OR "337513-53-8" OR "41318-75-6" OR "337513-56-1" OR "155999-95-4" OR "65075-08-3" OR "189084-60-4" OR "147217-78-5" OR "446254-17-7" OR "147217-80-9" OR "147217-79-6" OR "147217-81-0" OR "337513-54-9" OR "337513-68-5" OR "446254-18-8" OR "446254-19-9" OR "446254-20-2" OR "446254-22-4" OR "5436-43-1" OR "337513-55-0" OR "243982-82-3" OR "446254-23-5" OR "189084-57-9" OR "446254-24-6" OR "446254-25-7" OR "446254-31-5" OR "446254-32-6" OR "446254-33-7" OR "446254-34-8" OR "189084-61-5" OR "446254-37-1" OR "446254-38-2" OR "327185-09-1" OR "446254-39-3" OR "189084-62-6" OR "446254-40-6" OR "446254-41-7" OR "446254-42-8" OR "189084-63-7" OR "446254-43-9" OR "93703-48-1" OR "446254-45-1" OR "446254-48-4" OR "103173-66-6" OR "446254-50-8" OR "446254-51-9" OR "182346-21-0" OR "446254-53-1" OR "446254-54-2" OR "446254-55-3" OR "446254-55-3" OR "446254-57-5" OR "446254-59-7" OR "446254-61-1" OR "446254-64-4" OR "38463-82-0" OR "60348-60-9" OR "189084-64-8" OR "446254-65-5" OR "446254-66-6" OR "446254-67-7" OR "446254-68-8" OR "373594-78-6" OR "446254-69-9" OR "446254-71-3" OR "446254-72-4" OR "446254-74-6" OR "446254-77-9" OR "446254-78-0" OR "189084-65-9" OR "446254-80-4" OR "189084-66-0" OR "182677-30-1" OR "243982-83-4" OR "68631-49-2" OR "207122-15-4" OR "35854-94-5" OR "189084-58-0" OR "189084-67-1" OR "207122-16-5" OR "189084-68-2" OR "1163-19-5" OR "109945-70-2" OR "113152-37-7" OR "113172-79-5" OR "139598-16-6" OR "139749-52-3" OR "145538-74-5" OR "32534-81-9" OR "32536-52-0" OR "40088-47-9" OR "446254-27-9" OR "446255-20-5" OR "446255-22-7" OR "49690-94-0" OR "63936-56-1" OR "64589-00-0" OR "68928-80-3" OR "85446-17-9" OR "36483-60-0" OR "437701-79-6" OR "446255-26-1" OR "117948-63-7" OR "446255-30-7" OR "61262-53-1" OR "405237-85-6" OR "39275-89-3" OR "13654-09-6" OR "61288-13-9" OR "446255-39-6" OR "337513-72-1" OR "366791-32-4" OR "2050-47-7") AND ("Occupational Exposure"[Mesh:NoExp] OR "Maternal Exposure"[Mesh] OR "Environmental Exposure"[Mesh] OR "Prenatal Exposure Delayed Effects"[Mesh] OR "Exposure"[tw] OR "Exposed"[tw] OR "exposures"[tw] OR "exposing"[tw]) AND ("animal experimentation"[MeSH Terms] OR "models, animal"[MeSH Terms] OR "invertebrates"[MeSH Terms] OR "Animals"[Mesh:noexp] OR "animal population groups"[MeSH Terms] OR "mammals"[MeSH Terms:noexp] OR "primates"[MeSH Terms:noexp] OR "artiodactyla"[MeSH Terms] OR "carnivora"[MeSH Terms] OR "cetacea"[MeSH Terms] OR "chiroptera"[MeSH Terms] OR "elephants"[MeSH Terms] OR "hyraxes"[MeSH Terms] OR "insectivora"[MeSH Terms] OR "lagomorpha"[MeSH Terms] OR "marsupialia"[MeSH Terms] OR "monotremata"[MeSH Terms] OR "perissodactyla"[MeSH Terms] OR "rodentia"[MeSH Terms] OR "scandentia"[MeSH Terms] OR "sirenia"[MeSH Terms] OR "xenarthra"[MeSH Terms] OR "haplorhini"[MeSH Terms:noexp] OR "strepsirhini"[MeSH Terms] OR "platyrrhini"[MeSH Terms] OR "tarsii"[MeSH Terms] OR "catarrhini"[MeSH Terms:noexp] OR "cercopithecidae"[MeSH Terms] OR "hylobatidae"[MeSH Terms] OR "hominidae"[MeSH Terms:noexp] OR "gorilla gorilla"[MeSH Terms] OR "pan paniscus"[MeSH Terms] OR "pan troglodytes"[MeSH Terms] OR "pongo pygmaeus"[MeSH Terms] animals[tiab] OR animal[tiab] OR mice[Tiab] OR mus[Tiab] OR mouse[Tiab] OR murine[Tiab] OR woodmouse[tiab] OR rats[Tiab] OR rat[Tiab] OR murinae[Tiab] OR muridae[Tiab] OR cottonrat[tiab] OR cottonrats[tiab] OR hamster[tiab] OR hamsters[tiab] OR cricetinae[tiab] OR rodentia[Tiab] OR rodent[Tiab] OR rodents[Tiab] OR pigs[Tiab] OR pig[Tiab] OR swine[tiab] OR swines[tiab] OR piglets[tiab] OR piglet[tiab] OR boar[tiab] OR boars[tiab]

OR "sus scrofa"[tiab] OR ferrets[tiab] OR ferret[tiab] OR polecat[tiab] OR polecats[tiab] OR "mustela putorius"[tiab] OR "guinea pigs"[Tiab] OR "guinea pig"[Tiab] OR cavia[Tiab] OR callithrix[Tiab] OR marmoset[Tiab] OR marmosets[Tiab] OR cebuella[Tiab] OR hapale[Tiab] OR octodon[Tiab] OR chinchilla[Tiab] OR chinchillas[Tiab] OR gerbillinae[Tiab] OR gerbil[Tiab] OR gerbils[Tiab] OR jird[Tiab] OR jirds[Tiab] OR merione[Tiab] OR meriones[Tiab] OR rabbits[Tiab] OR rabbit[Tiab] OR hares[Tiab] OR hare[Tiab] OR cats[Tiab] OR cat[Tiab] OR felis[Tiab] OR dogs[Tiab] OR dog[Tiab] OR canine[Tiab] OR canines[Tiab] OR canis[Tiab] OR sheep[Tiab] OR sheeps[Tiab] OR mouflon[Tiab] OR mouflons[Tiab] OR ovis[Tiab] OR goats[Tiab] OR goat[Tiab] OR capra[Tiab] OR capras[Tiab] OR rupicapra[Tiab] OR chamois[Tiab] OR haplorhini[Tiab] OR monkey[Tiab] OR monkeys[Tiab] OR anthropoidea[Tiab] OR anthropoids[Tiab] OR saguinus[Tiab] OR tamarin[Tiab] OR tamarins[Tiab] OR leontopithecus[Tiab] OR hominidae[Tiab] OR ape[Tiab] OR apes[Tiab] OR pan[Tiab] OR paniscus[Tiab] OR "pan paniscus"[Tiab] OR bonobo[Tiab] OR bonobos[Tiab] OR "pan troglodytes"[Tiab] OR gibbon[Tiab] OR gibbons[Tiab] OR siamang[Tiab] OR siamangs[Tiab] OR nomascus[Tiab] OR symphalangus[Tiab] OR chimpanzee[Tiab] OR chimpanzees[Tiab] OR prosimians[Tiab] OR "bush baby"[Tiab] OR prosimian[Tiab] OR bush babies[Tiab] OR galagos[Tiab] OR galago[Tiab] OR pongidae[Tiab] OR gorilla[Tiab] OR gorillas[Tiab] OR pongo[Tiab] OR "pongo pygmaeus"[Tiab] OR orangutans[Tiab] OR lemur[Tiab] OR lemurs[Tiab] OR lemuridae[Tiab] OR horse[Tiab] OR horses[Tiab] OR pongo[Tiab] OR equus[Tiab] OR cow[Tiab] OR calf[Tiab] OR bull[Tiab] OR chicken[Tiab] OR chickens[Tiab] OR squirrel[Tiab] OR squirrels[Tiab] OR chipmunk[Tiab] OR chipmunks[Tiab] OR suslik[Tiab] OR susliks[Tiab] OR vole[Tiab] OR voles[Tiab] OR lemming[Tiab] OR lemmings[Tiab] OR muskrat[Tiab] OR muskrats[Tiab] OR lemmus[Tiab] OR otter[Tiab] OR otters[Tiab] OR marten[Tiab] OR martens[Tiab] OR martes[Tiab] OR weasel[Tiab] OR badger[Tiab] OR badgers[Tiab] OR ermine[Tiab] OR mink[Tiab] OR minks[Tiab] OR sable[Tiab] OR sables[Tiab] OR gulo[Tiab] OR gulos[Tiab] OR wolverine[Tiab] OR wolverines[Tiab] OR minks[Tiab] OR mustela[Tiab] OR llama[Tiab] OR llamas[Tiab] OR alpaca[Tiab] OR alpacas[Tiab] OR camelid[Tiab] OR camelids[Tiab] OR guanaco[Tiab] OR guanacos[Tiab] OR chiroptera[Tiab] OR chiropteras[Tiab] OR bat[Tiab] OR bats[Tiab] OR fox[Tiab] OR foxes[Tiab] OR donkey[Tiab] OR donkeys[Tiab] OR mule[Tiab] OR mules[Tiab] OR zebra[Tiab] OR zebras[Tiab] OR shrew[Tiab] OR shrews[Tiab] OR bison[Tiab] OR bisons[Tiab] OR buffalo[Tiab] OR buffaloes[Tiab] OR deer[Tiab] OR deers[Tiab] OR bear[Tiab] OR bears[Tiab] OR panda[Tiab] OR pandas[Tiab] OR "wild hog"[Tiab] OR "wild boar"[Tiab] OR fitchew[Tiab] OR fitch[Tiab] OR beaver[Tiab] OR beavers[Tiab] OR jerboa[Tiab] OR jerboas[Tiab] OR capybara[Tiab] OR capybaras[Tiab]) AND ("Attention"[Mesh] OR "attention"[tiab] OR "concentration"[tiab] OR "attentiveness"[tiab] OR "Behavior"[Mesh] OR "behavior"[tiab] OR "behaviour"[tiab] OR "behavioral"[tiab] OR "behavioural"[tiab] OR "behaviors"[tiab] OR "behaviours"[tiab] OR "Cognition"[Mesh] OR "Cognition Disorders"[Mesh] OR "cognition"[tiab] OR "cognitive"[tiab] OR "Developmental Disabilities"[Mesh] OR "developmental"[tiab] OR "Neurodevelopmental Disorders"[Mesh] OR "neurodevelopmental"[tiab] OR "neurodevelopment"[tiab] OR "neuropsychological"[tiab] OR "Executive Function"[Mesh] OR "executive function"[tiab] OR "executive functioning"[tiab] OR "Motor Activity"[Mesh] OR "locomotor"[tiab] OR "motor"[tiab] OR "Memory"[Mesh] OR "memory"[tiab] OR "Metacognition"[Mesh] OR "metacognition"[tiab] OR "metacognitive"[tiab] OR "Neurobehavioral Manifestations"[Mesh] OR "neurobehavioural"[tiab] OR "neurobehavioral"[tiab] OR "Neurotoxicity Syndromes"[Mesh] OR "neurotoxic"[tiab] OR "neurotoxicity" OR "neurotoxicant"[tiab] OR "neurotoxicants"[tiab] OR "neurotoxia"[tiab] OR "neurotoxicosis"[tiab] OR "processing speed"[tiab] OR "Spatial Learning"[Mesh] OR "spatial learning"[tiab] OR "Maze Learning"[Mesh] OR "maze"[tiab])

Embase

('flame retardant'/de OR '2,2`,4,4`,5,5` hexabromodiphenyl ether'/exp OR 'polybrominated diphenyl ether'/exp OR 'diphenyl ether derivative'/exp OR ((flame NEXT/1 retard*) OR (fire NEXT/1 retard*) OR (fireproofing NEXT/1 agent*) OR "FireMaster" OR "Bromkal" OR ('diphenyl ether' NEXT/1 deriv*) OR (Halogenated NEXT/1 Diphenyl*) OR (Brominated NEXT/1 Diphenyl*) OR PBDE* OR (Polybrominated NEXT/1 Diphenyl*) OR polybromodiphenyl* OR PBDP* OR BDE* OR pentabromo-

diphenyl* OR PentaBDE* OR "PeBDE" OR "DE 71" OR "DE71" OR "pentabrominated diphenyl" OR "pentabrominated diphenyls" OR "PBDPO" OR "Planelon PB 501" OR (pentabromo NEXT/1 deriv*) OR Pentabromophenyl* OR octabromodiphenyl* OR OctaBDE* OR "OcBDE" OR "Octabrom" OR "OBDE" OR "OBDPO" OR (octabromo NEXT/1 deriv*) OR "octabrominated diphenyl" OR "octabrominated diphenyls" OR decabromodiphenyl* OR DecaBDE* OR "DeBDE" OR "DBDPO" OR "decabrominated diphenyl" OR "decabrominated diphenyls" OR (decabromo NEXT/1 deriv*) OR "Decabrom" OR "Berkflam B 10E" OR "FR 300BA" OR "FR 300 BA" OR tribromodiphenyl* OR "tribrominated diphenyl" OR "tribrominated diphenyls" OR "TrBDE" OR (tribromo NEXT/1 deriv*) OR tetrabromodiphenyl* OR TetraBDE* OR "TeBDE" OR "TBDE" OR "BPDE" OR (tetrabromo NEXT/1 deriv*) OR "TBDP" OR "tetrabrominated diphenyl" OR "tetrabrominated diphenyls" OR hexabromodiphenyl* OR HexaBDE* OR "HxBDE" OR "hexabrominated diphenyl" OR "hexabrominated diphenyls" OR (hexabromo NEXT/1 deriv*) OR heptabromodiphenyl* OR HeptaBDE* OR "HeBDE" OR "heptabrominated diphenyl" OR "heptabrominated diphenyls" OR (heptabromo NEXT/1 deriv*) OR nonabromodiphenyl* OR NonaBDE* OR "NoBDE" OR "nonabrominated diphenyl" OR "nonabrominated diphenyls" OR (nonabromo NEXT/1 deriv*)):ti,ab,tn,rn OR ("7025-06-1" OR "6876-00-2" OR "101-55-3" OR "51452-87-0" OR "44625414-4" OR "147217-72-9" OR "171977-44-9" OR "147217-71-8" OR "33513-663" OR "51930-04-2" OR "6903-63-5" OR "189084-59-1" OR "83694-71-7" OR "46438-88-4" OR "2050-47-7" OR "147217-74-1" OR "147217-75-2" OR "407606-55-7" OR "147217-73-0" OR "147217-76-3" OR "337513-67-4" OR "446254-15-5" OR "446254-16-6" OR "147217-77-4" OR "337513-75-4" OR "337513-53-8" OR "41318-75-6" OR "337513-56-1" OR "155999-95-4" OR "65075-08-3" OR "189084-60-4" OR "147217-78-5" OR "446254-17-7" OR "147217-80-9" OR "147217-79-6" OR "147217-81-0" OR "337513-54-9" OR "337513-68-5" OR "446254-18-8" OR "446254-19-9" OR "446254-20-2" OR "446254-22-4" OR "5436-43-1" OR "337513-55-0" OR "243982-82-3" OR "446254-23-5" OR "189084-57-9" OR "446254-24-6" OR "446254-25-7" OR "446254-31-5" OR "446254-32-6" OR "446254-33-7" OR "446254-34-8" OR "189084-61-5" OR "446254-37-1" OR "446254-38-2" OR "327185-09-1" OR "446254-39-3" OR "189084-62-6" OR "446254-40-6" OR "446254-417" OR "446254-42-8" OR "189084-63-7" OR "446254-43-9" OR "93703-481" OR "446254-45-1" OR "446254-48-4" OR "103173-66-6" OR "446254-508" OR "446254-51-9" OR "182346-21-0" OR "446254-53-1" OR "446254-542" OR "446254-55-3" OR "446254-55-3" OR "446254-57-5" OR "44625459-7" OR "446254-61-1" OR "446254-64-4" OR "38463-82-0" OR "6034860-9" OR "189084-64-8" OR "446254-65-5" OR "446254-66-6" OR "446254-67-7" OR "446254-68-8" OR "373594-78-6" OR "446254-69-9" OR "446254-71-3" OR "446254-72-4" OR "446254-74-6" OR "446254-779" OR "446254-78-0" OR "189084-65-9" OR "446254-80-4" OR "18908466-0" OR "182677-30-1" OR "243982-83-4" OR "68631-49-2" OR "20712215-4" OR "35854-94-5" OR "189084-58-0" OR "189084-67-1" OR "20712216-5" OR "189084-68-2" OR "1163-19-5" OR "109945-70-2" OR "113152-37-7" OR "113172-79-5" OR "139598-16-6" OR "139749-52-3" OR "145538-74-5" OR "32534-81-9" OR "32536-52-0" OR "40088-47-9" OR "446254-27-9" OR "446255-20-5" OR "446255-22-7" OR "49690-94-0" OR "63936-56-1" OR "64589-00-0" OR "68928-80-3" OR "85446-17-9" OR "36483-60-0" OR "437701-79-6" OR "446255-26-1" OR "117948-63-7" OR "446255-30-7" OR "61262-53-1" OR "405237-85-6" OR "39275-89-3" OR "13654-09-6" OR "61288-13-9" OR "446255-39-6" OR "337513-72-1" OR "366791-32-4" OR "2050-47-7"):ti,ab,rn) AND ('ape'/de OR 'bat'/exp OR 'carnivora'/exp OR 'catarrhini'/de OR 'cercopithecidae'/exp OR 'cetacea'/exp OR 'chimpanzee'/exp OR 'chordata'/de OR 'elephant'/exp OR 'gorilla'/exp OR 'haplorhini'/de OR 'hominid'/de OR 'hylobatidae'/exp OR 'hyrax'/exp OR 'lagomorph'/exp OR 'mammal'/de OR 'marsupial'/exp OR 'monotremate'/exp OR 'orangutan'/exp OR 'placental mammals'/de OR 'platyrrhini'/exp OR 'primate'/de OR 'prosimian'/exp OR 'rodent'/exp OR 'scandentia'/exp OR 'simian'/de OR 'sirenia'/exp OR 'tarsiiform'/exp OR 'ungulate'/exp OR 'vertebrate'/de OR 'xenarthra'/exp OR animals:ti,ab OR animal:ti,ab OR mice:ti,ab OR mus:ti,ab OR mouse:ti,ab OR murine:ti,ab OR woodmouse:ti,ab OR rats:ti,ab OR rat:ti,ab OR murinae:ti,ab OR muridae:ti,ab OR cottonrat:ti,ab OR cottonrats:ti,ab OR hamster:ti,ab OR hamsters:ti,ab OR cricetinae:ti,ab OR rodentia:ti,ab OR rodent:ti,ab OR rodents:ti,ab OR pigs:ti,ab OR pig:ti,ab OR swine:ti,ab OR

swines:ti,ab OR piglets:ti,ab OR piglet:ti,ab OR boar:ti,ab OR boars:ti,ab OR "sus scrofa":ti,ab OR ferrets:ti,ab OR ferret:ti,ab OR polecat:ti,ab OR polecats:ti,ab OR "mustela putorius":ti,ab OR "guinea pigs":ti,ab OR "guinea pig":ti,ab OR cavia:ti,ab OR callithrix:ti,ab OR marmoset:ti,ab OR marmosets:ti,ab OR cebuella:ti,ab OR hapale:ti,ab OR octodon:ti,ab OR chinchilla:ti,ab OR chinchillas:ti,ab OR gerbillinae:ti,ab OR gerbil:ti,ab OR gerbils:ti,ab OR jird:ti,ab OR jirds:ti,ab OR merione:ti,ab OR meriones:ti,ab OR rabbits:ti,ab OR rabbit:ti,ab OR hares:ti,ab OR hare:ti,ab OR cats:ti,ab OR cat:ti,ab OR felis:ti,ab OR dogs:ti,ab OR dog:ti,ab OR canine:ti,ab OR canines:ti,ab OR canis:ti,ab OR sheep:ti,ab OR sheeps:ti,ab OR mouflon:ti,ab OR mouflons:ti,ab OR ovis:ti,ab OR goats:ti,ab OR goat:ti,ab OR capra:ti,ab OR capras:ti,ab OR rupicapra:ti,ab OR chamois:ti,ab OR haplorhini:ti,ab OR monkey:ti,ab OR monkeys:ti,ab OR anthropoidea:ti,ab OR anthropoids:ti,ab OR saguinus:ti,ab OR tamarin:ti,ab OR tamarins:ti,ab OR leontopithecus:ti,ab OR hominidae:ti,ab OR ape:ti,ab OR apes:ti,ab OR pan:ti,ab OR paniscus:ti,ab OR "pan paniscus":ti,ab OR bonobo:ti,ab OR bonobos:ti,ab OR "pan troglodytes":ti,ab OR gibbon:ti,ab OR gibbons:ti,ab OR siamang:ti,ab OR siamangs:ti,ab OR nomascus:ti,ab OR symphalangus:ti,ab OR chimpanzee:ti,ab OR chimpanzees:ti,ab OR prosimians:ti,ab OR "bush baby":ti,ab OR prosimian:ti,ab OR bush babies:ti,ab OR galagos:ti,ab OR galago:ti,ab OR pongidae:ti,ab OR gorilla:ti,ab OR gorillas:ti,ab OR pongo:ti,ab OR "pongo pygmaeus":ti,ab OR orangutans:ti,ab OR lemur:ti,ab OR lemurs:ti,ab OR lemuridae:ti,ab OR horse:ti,ab OR horses:ti,ab OR pongo:ti,ab OR equus:ti,ab OR cow:ti,ab OR calf:ti,ab OR bull:ti,ab OR chicken:ti,ab OR chickens:ti,ab OR squirrel:ti,ab OR squirrels:ti,ab OR chipmunk:ti,ab OR chipmunks:ti,ab OR suslik:ti,ab OR susliks:ti,ab OR vole:ti,ab OR voles:ti,ab OR lemming:ti,ab OR lemmings:ti,ab OR muskrat:ti,ab OR muskrats:ti,ab OR lemmus:ti,ab OR otter:ti,ab OR otters:ti,ab OR marten:ti,ab OR martens:ti,ab OR martes:ti,ab OR weasel:ti,ab OR badger:ti,ab OR badgers:ti,ab OR ermine:ti,ab OR mink:ti,ab OR minks:ti,ab OR sable:ti,ab OR sables:ti,ab OR gulo:ti,ab OR gulos:ti,ab OR wolverine:ti,ab OR wolverines:ti,ab OR minks:ti,ab OR mustela:ti,ab OR llama:ti,ab OR llamas:ti,ab OR alpaca:ti,ab OR alpacas:ti,ab OR camelid:ti,ab OR camelids:ti,ab OR guanaco:ti,ab OR guanacos:ti,ab OR chiroptera:ti,ab OR chiropteras:ti,ab OR bat:ti,ab OR bats:ti,ab OR fox:ti,ab OR foxes:ti,ab OR donkey:ti,ab OR donkeys:ti,ab OR mule:ti,ab OR mules:ti,ab OR zebra:ti,ab OR zebras:ti,ab OR shrew:ti,ab OR shrews:ti,ab OR bison:ti,ab OR bisons:ti,ab OR buffalo:ti,ab OR buffaloes:ti,ab OR deer:ti,ab OR deers:ti,ab OR bear:ti,ab OR bears:ti,ab OR panda:ti,ab OR pandas:ti,ab OR "wild hog":ti,ab OR "wild boar":ti,ab OR fitchew:ti,ab OR fitch:ti,ab OR beaver:ti,ab OR beavers:ti,ab OR jerboa:ti,ab OR jerboas:ti,ab OR capybara:ti,ab OR capybaras:ti,ab) AND ("attention"/exp OR "attention":ti,ab OR "concentration":ti,ab OR "attentiveness":ti,ab OR "behavior"/exp OR "behavior":ti,ab OR "behaviour":ti,ab OR "behavioral":ti,ab OR "behavioural":ti,ab OR "behaviors":ti,ab OR "behaviours":ti,ab OR "cognition"/exp OR "cognition":ti,ab OR "cognitive":ti,ab OR "cognition assessment"/exp OR "developmental disorder"/exp OR "developmental":ti,ab OR "executive function"/exp OR "executive function":ti,ab OR "executive functioning":ti,ab OR "motor activity"/exp OR "locomotor":ti,ab OR "motor":ti,ab OR "memory"/exp OR "memory":ti,ab OR "metacognition"/exp OR "metacognition":ti,ab OR "metacognitive":ti,ab OR "neurobehavioural":ti,ab OR "neurobehavrioral":ti,ab OR "neurotoxicity"/exp OR "neurotoxic":ti,ab OR "neurotoxicity" OR "neurotoxicant":ti,ab OR "neurotoxicants":ti,ab OR "neurotoxia":ti,ab OR "neurotoxicosis":ti,ab OR "processing speed":ti,ab OR "spatial learning"/exp OR "spatial learning":ti,ab OR "maze test"/exp OR "maze":ti,ab)

Toxline

("flame retard*" OR "fire retard*" OR "fireproofing agent*" OR "FireMaster" OR "Bromkal" OR "diphenyl ether deriv*" OR "Halogenated Diphenyl*" OR "Brominated Diphenyl*" OR PBDE* OR "Polybrominated Diphenyl*" OR polybromodiphenyl* OR PBDP* OR BDE* OR pentabromodiphenyl* OR "c-pentaBDE*" OR PentaBDE* OR "PeBDE" OR "DE 71" OR "DE71" OR "pentabrominated diphenyl*" OR "PBDPO" OR "Planelon PB 501" OR "pentabromo deriv*" OR Pentabromophenyl* OR octabromodiphenyl* OR "c-octaBDE*" OR OctaBDE* OR "OcBDE" OR "Octabrom" OR "octabromo deriv*" OR "OBDE" OR "OBDPO" OR "octabrominated diphenyl*" OR decabromodiphenyl* OR "cdecaBDE*" OR DecaBDE* OR "DeBDE" OR "DBDPO" OR "decabrominated diphenyl*" OR

"decabromo deriv*" OR "Decabrom" OR "Berkflam B 10E" OR "FR 300BA" OR "FR 300 BA" OR tribromodiphenyl* OR "tribrominated diphenyl*" OR "TrBDE" OR "tribromo deriv*" OR tetrabromodiphenyl* OR TetraBDE* OR "TeBDE" OR "TBDE" OR "BPDE" OR "tetrabromo deriv*" OR "TBDP" OR "tetrabrominated diphenyl*" OR hexabromodiphenyl* OR HexaBDE* OR "HxBDE" OR "hexabrominated diphenyl*" OR "hexabromo deriv*" OR heptabromodiphenyl* OR HeptaBDE* OR "HeBDE" OR "heptabrominated diphenyl*" OR "heptabromo deriv*" OR nonabromodiphenyl* OR NonaBDE* OR "NoBDE" OR "nonabrominated diphenyl*" OR "nonabromo deriv*" OR "7025-06-1" OR "6876-00-2" OR "101-55-3" OR "51452-87-0" OR "446254-14-4" OR "147217-72-9" OR "171977-44-9" OR "147217-71-8" OR "33513-66-3" OR "51930-04-2" OR "6903-63-5" OR "189084-59-1" OR "83694-71-7" OR "46438-88-4" OR "2050-47-7" OR "147217-74-1" OR "147217-75-2" OR "407606-55-7" OR "147217-73-0" OR "147217-763" OR "337513-67-4" OR "446254-15-5" OR "446254-16-6" OR "14721777-4" OR "337513-75-4" OR "337513-53-8" OR "41318-75-6" OR "337513-56-1" OR "155999-95-4" OR "65075-08-3" OR "189084-60-4" OR "147217-78-5" OR "446254-17-7" OR "147217-80-9" OR "147217-796" OR "147217-81-0" OR "337513-54-9" OR "337513-68-5" OR "44625418-8" OR "446254-19-9" OR "446254-20-2" OR "446254-22-4" OR "5436-43-1" OR "337513-55-0" OR "243982-82-3" OR "446254-23-5" OR "189084-57-9" OR "446254-24-6" OR "446254-25-7" OR "446254-31-5" OR "446254-32-6" OR "446254-33-7" OR "446254-348" OR "189084-61-5" OR "446254-37-1" OR "446254-38-2" OR "327185-09-1" OR "446254-39-3" OR "189084-62-6" OR "446254-406" OR "446254-41-7" OR "446254-42-8" OR "189084-63-7" OR "446254-43-9" OR "93703-48-1" OR "446254-45-1" OR "446254-48-4" OR "103173-66-6" OR "446254-50-8" OR "446254-51-9" OR "18234621-0" OR "446254-53-1" OR "446254-54-2" OR "446254-55-3" OR "446254-55-3" OR "446254-57-5" OR "446254-59-7" OR "446254-611" OR "446254-64-4" OR "38463-82-0" OR "60348-60-9" OR "189084-64-8" OR "446254-65-5" OR "446254-66-6" OR "446254-677" OR "446254-68-8" OR "373594-78-6" OR "446254-69-9" OR "446254-71-3" OR "446254-72-4" OR "446254-74-6" OR "446254-779" OR "446254-78-0" OR "189084-65-9" OR "446254-80-4" OR "189084-66-0" OR "182677-30-1" OR "243982-83-4" OR "68631-49-2" OR "207122-15-4" OR "35854-94-5" OR "189084-58-0" OR "18908467-1" OR "207122-16-5" OR "189084-68-2" OR "1163-19-5" OR "109945-70-2" OR "113152-37-7" OR "113172-79-5" OR "139598-16-6" OR "139749-52-3" OR "145538-74-5" OR "32534-81-9" OR "32536-520" OR "40088-47-9" OR "446254-27-9" OR "446255-20-5" OR "446255-22-7" OR "49690-94-0" OR "63936-56-1" OR "64589-00-0" OR "68928-80-3" OR "85446-17-9" OR "36483-60-0" OR "437701-796" OR "446255-26-1" OR "117948-63-7" OR "446255-30-7" OR "6126253-1" OR "405237-85-6" OR "39275-89-3" OR "13654-09-6" OR "61288-13-9" OR "446255-39-6" OR "337513-72-1" OR "366791-32-4" OR "2050-47-7") AND (animals OR animal OR mice OR mus OR mouse OR murine OR woodmouse OR rats OR rat OR murinae OR muridae OR cottonrat OR cottonrats OR hamster OR hamsters OR cricetinae OR rodentia OR rodent OR rodents OR pigs OR pig OR swine OR swines OR piglets OR piglet OR boar OR boars OR "sus scrofa" OR ferrets OR ferret OR polecat OR polecats OR "mustela putorius" OR "guinea pigs" OR "guinea pig" OR cavia OR callithrix OR marmoset OR marmosets OR cebuella OR hapale OR octodon OR chinchilla OR chinchillas OR gerbillinae OR gerbil OR gerbils OR jird OR jirds OR merione OR meriones OR rabbits OR rabbit OR hares OR hare OR cats OR cat OR felis OR dogs OR dog OR canine OR canines OR canis OR sheep OR sheeps OR mouflon OR mouflons OR ovis OR goats OR goat OR capra OR capras OR rupicapra OR chamois OR haplorhini OR monkey OR monkeys OR anthropoidea OR anthropoids OR saguinus OR tamarin OR tamarins OR leontopithecus OR hominidae OR ape OR apes OR pan OR paniscus OR "pan paniscus" OR bonobo OR bonobos OR "pan troglodytes" OR gibbon OR gibbons OR siamang OR siamangs OR nomascus OR symphalangus OR chimpanzee OR chimpanzees OR prosimians OR "bush baby" OR prosimian OR bush babies OR galagos OR galago OR pongidae OR gorilla OR gorillas OR pongo OR "pongo pygmaeus" OR orangutans OR lemur OR lemurs OR lemuridae OR horse OR horses OR pongo OR equus OR cow OR calf OR bull OR chicken OR chickens OR squirrel OR squirrels OR chipmunk OR chipmunks OR suslik OR susliks OR vole OR voles OR lemming OR lemmings OR muskrat OR muskrats OR lemmus OR otter OR otters OR marten OR martens OR martes OR weasel OR

badger OR badgers OR ermine OR mink OR minks OR sable OR sables OR gulo OR gulos OR wolverine OR wolverines OR minks OR mustela OR llama OR llamas OR alpaca OR alpacas OR camelid OR camelids OR guanaco OR guanacos OR chiroptera OR chiropteras OR bat OR bats OR fox OR foxes OR donkey OR donkeys OR mule OR mules OR zebra OR zebras OR shrew OR shrews OR bison OR bisons OR buffalo OR buffaloes OR deer OR deers OR bear OR bears OR panda OR pandas OR "wild hog" OR "wild boar" OR fitchew OR fitch OR beaver OR beavers OR jerboa OR jerboas OR capybara OR capybaras) AND ("Exposure" OR "Exposed" OR "exposures" OR "exposing") AND ("attention" OR "concentration" OR "attentiveness" OR "behavior" OR "behaviour" OR "behavioral" OR "behavioural" OR "behaviors" OR "behaviours" OR "Cognition Disorders"[Mesh] OR "cognition" OR "cognitive" OR "developmental" OR "executive function" OR "executive functioning" OR "locomotor" OR "motor" OR "memory" OR "metacognition" OR "metacognitive" OR "neurobehavioural" OR "neurobehavrioral" OR "neurotoxic" OR "neurotoxicity" OR "neurotoxicant" OR "neurotoxicants" OR "neurotoxia" OR "neurotoxicosis" OR "processing speed" OR "spatial learning" OR "maze")

SECTION E-1c

SCREENING FORMS

Title and Abstract Screening Form

Instructions: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include nonhuman mammals	
Study does not report PBDE exposure	
No relevant outcomes	
Incomplete information (e.g., conference abstract, meeting poster)	
Not in English and unable to determine eligibility	
Other (explanation required)	

Full-Text Screening Form

Instructions: When a citation is excluded, reason should be specified.

Exclusion Reasons	
No original data (e.g., review article, commentary, editorial)	
Study does not include nonhuman mammals	
Study does not report PBDE exposure	
Study does not quantify exposure to PBDE	
Study does not include developmental exposure	
Study does not assess or report quantitative measures of learning, memory, attention, or response inhibition	
No comparator group (different doses or vehicle-only treatment)	
Not in English and unable to determine eligibility	
Other (explanation required)	

SECTION E-1d

DATA EXTRACTION ELEMENTS FOR ANIMAL STUDIES

Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or life stage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls—was expected response observed? (*information bias)
	End point health category (e.g., reproductive)
	End point (e.g., infertility)
	Diagnostic or method to measure end point (*information bias)
	Statistical methods (*information bias)
Results	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. When possible, measures of effect will be converted to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
	No-observed-effect level (NOEL), lowest-observed-effect level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. Note: The NOEL and LOEL 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 NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, nonmonotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias.

SECTION E-1e

RISK OF BIAS QUESTIONS FOR ANIMAL STUDIES

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
• Direct evidence that animals were allocated to any study group including controls using a method with a random component, • **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups. • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green, 2011). • **Note:** Restricted randomization (e.g., blocked randomization) to ensure that particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
• Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that animals were allocated to study groups using a method with a nonrandom component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how animals were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that animals were allocated to study groups using a nonrandom method, including judgment of the investigator, the results of a laboratory test, or a series of tests, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
• Direct evidence that the same vehicle was used in control and experimental animals, • **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that the same vehicle was used in control and experimental animals, • **OR** it is deemed that the vehicle used would not appreciably bias results, • **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the vehicle differed between control and experimental animals, • **OR** authors did not report the vehicle used (record "NR" as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
• Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than were experimental animals, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of the study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, • **Note:** Acceptable handling of attrition includes very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or 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; missing outcomes is not enough to impact the effect estimate. • **OR** missing data have been imputed using appropriate methods (ensuring that characteristics of animals are not significantly different from animals retained in the analysis).
Probably Low Risk of Bias (+)
• Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, • **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of animals was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of animals was unacceptably large and not adequately addressed. • **Note:** Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that the exposure to the PBDE was independently characterized (including purity, stability, and compliance with the treatment, if applicable) and confirmed generally as ≥98% purity, • **OR** direct evidence that all individual congeners were independently assessed for purity if a "mixture" is developed by the researchers, • **OR** the mixture should be independently assessed and non-target congeners or other impurities confirmed to contribute less than 2% (purity is ≥98%), • **AND** that exposure was consistently administered (i.e., with the same method and time frame) across treatment groups, • **AND** for gavage, dietary, or drinking water studies, that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups, • **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure to the PBDE was independently characterized (including purity, stability, and compliance with the treatment, if applicable) and confirmed generally as ≥98% (i.e., the supplier of the chemical provides documentation of the purity of the chemical), • **OR** indirect evidence that all individual congeners were independently assessed for purity if a "mixture" is developed by the researchers (the supplier of the chemical provides documentation of the purity of each chemical) and non-target congeners/impurities confirmed as less than 98%, • **OR** the mixture is provided by a supplier and the supplier provides documentation of the purity of the mixture with non-target congeners/impurities confirmed to contribute less than 2% (i.e. purity is ≥98%), • **OR** direct evidence that the purity of the congener(s) was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results, • **AND** that exposure was consistently administered (i.e., with the same method and time frame) across treatment groups, • **AND** for dietary or drinking water studies, no information is provided on consumption or internal dose metrics,

• **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record "NR" as basis for answer), • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
• Direct evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., Morris water maze, radial arm maze, operant tests of cognition) • **AND** assessed at the same length of time (i.e., same day of life) after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes.
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable), • **AND** assessed at the same length of time (i.e., same day of life) after initial exposure in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, 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, which is more likely to apply to objective outcome measures.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported.

Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes 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 in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes 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 a composite score without individual outcome components or outcomes reported using measurements, analysis methods, or subsets of the data (e.g., subscales) that were not prespecified or reporting outcomes not prespecified, or that unplanned analyses were included that would appreciably bias results.

11. Was litter or litter effects considered appropriately in the statistical analyses and were there no other potential threats to internal validity?

Because this evaluation is focused on developmental exposure, this question was added to address litter effects in data analysis. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk of bias considerations that do not fit under the other questions.

Definitely Low Risk of Bias (++)
• Direct evidence that litter effects were appropriately considered in the study design or analysis, using one of the following approaches: • The dam used as the statistical unit of analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects were appropriately considered in the analysis AND the statistical method was stated.
Probably Low Risk of Bias (+)
• Indirect evidence that litter effects were appropriately considered in the study design or analysis, using one of the following approaches: • The dam used as the statistical unit of analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter-effects were appropriately considered in the analysis BUT the statistic method used to address litter effects was not stated.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that litter effects were not appropriately considered in the study design or analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter-effects **were not** considered in the statistical analysis.
Definitely High Risk of Bias (--)
• Direct evidence that litter effects were not appropriately considered in the study design or analysis, • **OR** the fetus/pup used as the statistical unit of analysis AND litter effects **were not** considered in the statistical analysis.

SECTION E-1f

AMENDMENTS TO THE PROTOCOL

Changes to the Review Team (September 15, 2016)

Original review team members Barbara Hales and Susan Schantz were replaced by the following committee members who have more experience conducting risk of bias evaluations and data extraction:

- **David C. Dorman** (*Chair*) is a professor of toxicology in the Department of Molecular Biosciences of North Carolina State University. The primary objective of his research is to provide a refined understanding of chemically induced neurotoxicity in laboratory animals that will lead to improved assessment of potential toxicity in humans. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has chaired or served on several NRC committees, including the Committee on Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decisions, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He has served on other advisory boards for the US Navy, NASA, and USDA, and is currently a member of NTP's Board of Scientific Counselors. Dr. Dorman is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Sciences. He received a DVM from Colorado State University. He completed a combined PhD and residency program in toxicology at the University of Illinois at Urbana-Champaign, and he is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.
- **Andrew A. Rooney** is deputy director of the Office of Health Assessment and Translation (OHAT) in the National Toxicology Program at the National Institute of Environmental Health Sciences. He has been developing risk assessment methods and guidance throughout his professional career and is a principal author of the 2012 WHO/IPCS Guidance for Immunotoxicity Risk Assessment for Chemicals. Most recently, he has been working on emerging issues in toxicology and environmental health, including methods to address study quality in terms of risk of bias for human, animal, and mechanistic studies and adaptation of systematic review methods for addressing environmental health questions. He led the team that developed the OHAT approach to systematic review. Dr. Rooney has an MS and a PhD in zoology from the University of Florida.

SECTION E-2

Results of Literature Searches for Animal Studies on the Effects of Developmental Exposure to PBDEs on Learning, Memory, Attention, or Response Inhibition

Literature searches were performed on August 15, 2016, using the search strategy presented in the PBDE (Animal) Systematic Review Protocol (Section E-1). A summary of the results is presented below.

Embase:	1,326
PubMed:	1,173
Toxline:	489

Total citations found:	2,988
Duplicates removed:	1,137
Total unique citations:	1,851

SECTION E-3

Funding Sources of the Animal Studies on PBDEs and Learning, Memory, or Attention

Sources of funding were used to evaluate publication bias in terms of whether a particular sector funded more studies than another.

Reference	Governmental NIH/Federal	Industry	Other	Unknown
Biesemeier et al. 2011		X		
Blanco et al. 2013	X (Spain)			
Bowers et al. 2015	X (Canada)			
Buratovic et al. 2014	X (Sweden)			
Chen et al. 2014	X (China)			
Cheng et al. 2009	X (China)			
de-Miranda et al. 2016	X (Brazil)			
Driscoll et al. 2009	X (Colorado)			
Driscoll et al. 2012				X
Dufault et al. 2005	X (Colorado)			
Eriksson et al. 2001	X (Sweden)			
Fischer et al. 2008	X (Sweden)			
He et al. 2009	X (China)			
He et al. 2011	X (China)			
Koenig et al. 2012	X (NIEHS)		JB Johnson Foundation	
Llansola et al. 2009	X (EU)			
Reverte et al. 2013	X (FEDR; EU)			
Reverte et al. 2014	X (FEDR; EU)			
Rice et al. 2009	X (Maine)			
Ta et al. 2011	X (NIEHS; EPA)			
Verma et al. 2013	X (India)			
Verma et al. 2014	X (India)			
Viberg et al. 2003	X (Sweden; EU)			
Viberg et al. 2006	X (EU)			
Woods et al. 2012	X (NIH; NIEHS; EPA)			
Zhang et al. 2013	X (China)			
Zhao et al. 2014	X (China)			

SECTION E-4

Confidence Ratings for the Body of Evidence from Animal Studies of PBDEs

The body of evidence from animal studies of the PBDEs and learning, memory, and attention, were rated in accordance with the guidance presented in Section E-1. No studies of response inhibition were found.

BDE-47

Studies of BDE-47 and effects on learning (see Table E4-1) and memory (see Table E4-2) were available.

Learning: There is moderate confidence in the body of evidence on developmental exposure to BDE-47 and effects on learning in rodents. Six studies in mice and rats were available. The two studies in rats found several indications of decreased learning in the Morris water maze (e.g., prolonged latency periods) after treatment with BDE-47 at doses of 1, 5, or 10 mg/kg-day on PND 10. Both studies were from the same laboratory (He et al. 2009, 2011). Three of the four mouse studies reported decreased learning in at least one test, strain, or sex and were conducted by different research groups (Eriksson et al. 2001; Ta et al. 2011; Koenig et al. 2012; Woods et al. 2012). Nevertheless, the mouse results were variable across all the tests administered and a clear pattern was not identified to explain the heterogeneity in response relative to a susceptible strain, sex, or dose.

- **Risk of bias:** Downgraded because all studies had at least one rating of probably high or definitely high risk of bias in one of the key issues (e.g., lack of randomization of treatment), and most of the studies had multiple risk of bias issues, including not controlling for litter effects in the study design or analysis (see Figure E4-1).
- **Unexplained inconsistencies:** A qualitative evaluation of the evidence suggested a possible downgrade because of the heterogeneity in the evidence. However, a meta-analysis of studies of several PBDEs (see Chapter 4 and Appendix E, Section E-5), including BDE-47, and latency in the last trial of the Morris water maze showed consistent evidence of an effect on this measure of learning, so confidence was not downgraded.

TABLE E4-1 Studies of BDE-47 and Learning in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
He et al. 2009	Sprague-Dawley rats	PND 10	2 months	Morris water maze	None	1
He et al. 2011	Sprague-Dawley rats	PND 10	2 months	Morris water maze	None	1
Koenig et al. 2012	C57BL/6J mice	GD 0 - PND 21	2 months	Barnes maze	None	0.03
Ta et al. 2011	C57BL/6J mice	GD 0 - PND 21	2 months	Morris water maze	0.1	1
Woods et al. 2012	Female Mecp2 308+/- mice	GD 0 - PND 21	PND 50-54	Morris water maze	None	0.03
	Male Mecp2 308+/- mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Female C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Male C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

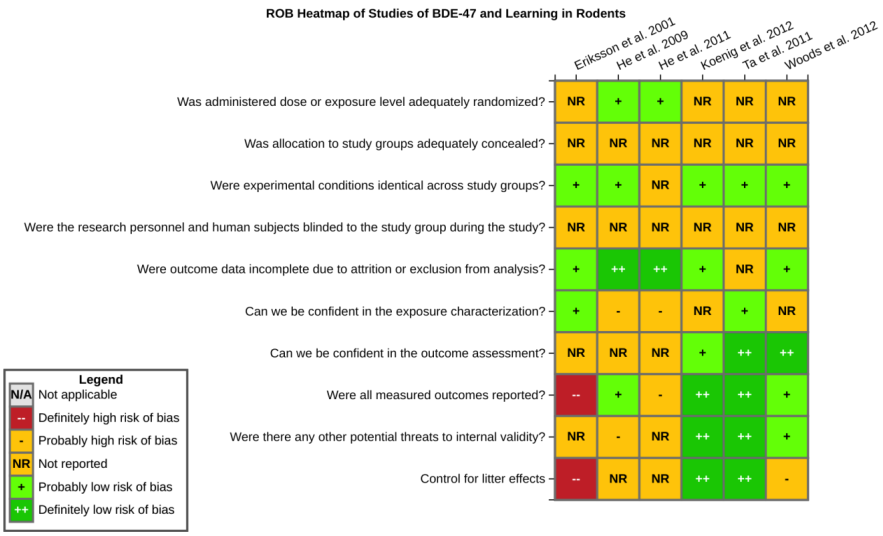


FIGURE E4-1 Risk of bias heatmap of studies of BDE-47 and learning in rodents. In HAWC: https://hawcproject.org/summary/visual/353/.

- **Indirectness:** No downgrade because tests used are considered direct measures of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade. Only two studies in rats were available, and the studies were from the same laboratory. The evidence base was judged to be inadequate for making judgments about cross-species consistency.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

Memory: There is low confidence in the body of evidence on developmental exposure to BDE-47 and effects on memory in rodents. Five studies in rats and mice were available. The one study in rats (He et al. 2011) reported decreased memory in the Morris water maze (e.g., prolonged latency periods) after exposure at 1, 5, and 10 mg/kg-day on PND 10. Three of the mouse studies (Eriksson et al. 2001; Ta et al. 2011; Koenig et al. 2012) reported no effects on memory at doses up to 12 mg/kg-day; however, Woods et al. (2012) reported decrements in memory in female Mecp2 308+/– mice with no effects on males or in C57BL6 mice of either sex. The body of evidence contained only a single study in rats and inconsistent results in mice. The dataset is similar and contains some of the same studies discussed above with respect to effects of BDE-47 on learning; fewer studies reported an effect, however, and one less study overall.

- **Risk of bias:** Downgraded because all the studies had a probably high risk of bias rating for at least one major issue (e.g., researchers were not blinded to the study groups during outcome assessment), and most of the studies had multiple risk of bias issues, including not controlling for litter effects in the study design or analysis (Eriksson et al. 2001; Woods et al. 2012) (see Figure E4-2).

TABLE E4-2 Studies of BDE-47 and Memory in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
He et al. 2009	Sprague-Dawley rats	PND 10	2 months	Morris water maze	None	1
Koenig et al. 2012	C57BL/6J mice	GD 0 - PND 21	2 months	Barnes maze	1	None
Ta et al. 2011	C57BL/6J mice	GD 0 - PND 21	2 months	Morris water maze	1	None
Woods et al. 2012	Female Mecp2 308+/- mice	GD 0 - PND 21	PND 50-54	Morris water maze	None	0.03
	Male Mecp2 308+/- mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Female C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None
	Male C57Bl6 mice	GD 0 - PND 21	PND 50-54	Morris water maze	0.03	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

- **Unexplained inconsistencies:** Downgraded because of the heterogeneity of the evidence.
- **Indirectness:** No downgrade because tests used are considered direct measures of memory.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade. Only one study in rats was available, so the evidence base was judged to be inadequate to make judgments about cross-species consistency.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

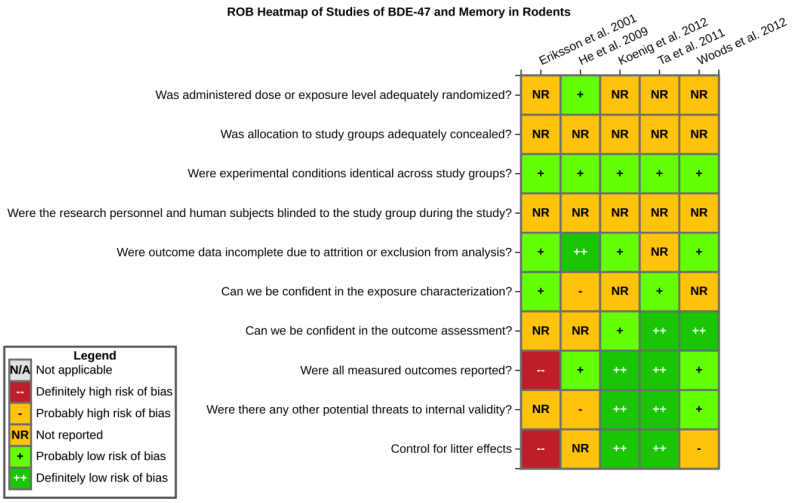


FIGURE E4-2 Risk of bias heatmap of studies of BDE-47 and memory in rodents. In HAWC: https://hawcproject.org/summary/visual/354/.

BDE-99

Studies of BDE-99 and effects on learning (see Table E4-3) and memory (see Table E4-4) were available.

TABLE E4-3 Studies of BDE-99 and Learning in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Blanco et al. 2013	Sprague-Dawley rats	GD 6 to PND 21	PND 26-35	Morris water maze	1	2
Cheng et al. 2009	Sprague-Dawley rats	GD 6 - PND 21	PND 34-36	Morris water maze	None	2
Fischer et al. 2008	NMRI mice	PND 10	4 months	Morris water maze	None	0.8
	NMRI mice	PND 10	4 months	Radial maze	None	0.8
Llansola et al. 2009	Wistar rats	GD 2-9	PND 68-70	Y maze	30	None
	Wistar rats	GD 11-19	PND 68-70	Y maze	30	None
Zhao et al. 2014	Sprague-Dawley rats	GD 1 - PND 21	PND 34-36	Morris water maze	0.2	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Learning: There is moderate confidence in the body of evidence on developmental exposure to BDE-99 and learning in mice and rats based on five studies. Two of the three studies in rats reported longer latency during the acquisition period of tests using the Morris water maze at a dose of 2 mg/kg-day (Cheng et al. 2009; Blanco et al. 2013). Zhao et al. (2014) reported no effects at a lower dose (0.2 mg/kg-day) under similar exposure and testing conditions. In contrast, developmental exposure of Wistar rats at doses up to 30 mg/kg-day had no effect on learning tested with a Y maze (Llansola et al. 2009). A single study (Fischer et al. 2008) in NMRI mice also reported decrements in learning during the acquisition period in tests using either a radial maze or a Morris water maze at a dose of 0.8 mg/kg-day.

- **Risk of bias:** Downgraded because of serious concerns about several risk of bias issues. All of the studies were rated as having probably high risk of bias for at least one key risk of bias issue (e.g., researchers were not blinded to the study groups during outcome assessment), two of the studies had a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis, and most of the studies had multiple risk of bias issues (see Figure E4-3).
- **Unexplained inconsistencies:** A qualitative evaluation of the evidence suggested a possible downgrade because of the heterogeneity in the evidence. Nevertheless, a meta-analysis of studies of several PBDEs (see Chapter 4 and Appendix E, Section E-5), including BDE-99, and latency in the last trial of the Morris water maze showed consistent evidence of an effect on this measure of learning, so confidence was not downgraded.
- **Indirectness:** No downgrade because tests used were considered direct measures of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

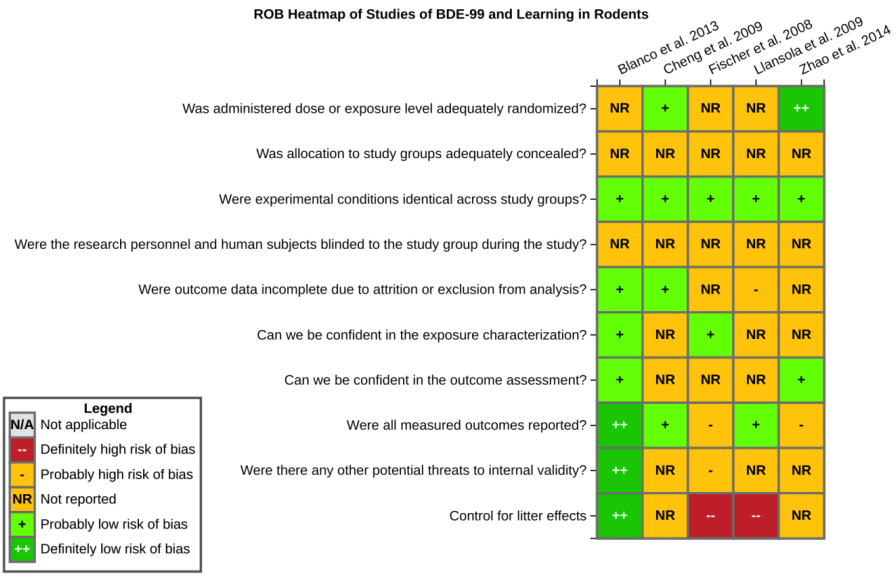


FIGURE E4-3 Risk of bias heatmap of studies of BDE-99 and learning in rodents. In HAWC: https://hawcproject.org/summary/visual/355/.

TABLE E4-4 Studies of BDE-99 and Memory in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Blanco et al. 2013	Sprague-Dawley rats	GD 6 - PND 21	PND 26-35	Morris water maze	2	None
Fischer et al. 2008	NMRI mice	PND 10	6 months	Radial maze	0.8	None
Zhao et al. 2014	Sprague-Dawley rats	GD 1 - PND 21	PND 34-36	Morris water maze	0.2	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Memory: There is moderate confidence in the data to evaluate whether developmental exposure to BDE-99 affects memory in rodents. The three available studies found no effects in several memory tests at doses of 0.2-2 mg/kg-day.

- **Risk of bias:** Downgraded because of serious concerns about several risk of bias issues. All the studies had a rating of probably high risk of bias for at least one key risk of bias issue (e.g., researchers were not blinded to the study groups during outcome assessment) and one study had a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis (see Figure E4-4).
- **Unexplained inconsistencies:** No downgrade for inconsistency.
- **Indirectness:** No downgrade because tests used are considered direct measures of memory.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because one study in mice and two in rats is insufficient to upgrade for evidence of consistency across species for a *no-effect* finding.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

FIGURE E4-4 Risk of bias heatmap of studies of BDE-99 and memory in rodents. In HAWC: https://hawcproject.org/summary/visual/356/.

BDE-153

Studies of BDE-153 and effects on learning (see Table E4-5) and memory (see Table E4-6) were available.

TABLE E4-5 Studies of BDE-153 and Learning in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Viberg et al. 2003	NMRI mouse	PND 10	PND 180	Morris water maze	0.45	0.9
Zhang et al. 2013	Sprague-Dawley rats	PND 10	PND 40	Morris water maze	10	None
	Sprague-Dawley rats	PND 10	PND 70	Morris water maze	10	None
	Sprague-Dawley rats	PND 10	PND 40	Passive avoidance	10	None
	Sprague-Dawley rats	PND 10	PND 70	Passive avoidance	10	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Learning: There is low confidence in the body of evidence to evaluate whether developmental exposure to BDE-153 affects learning in mice or rats. Two studies were available, one in mice and one in rats. Viberg et al. (2003) reported longer latencies in the acquisition period in the Morris water maze when mice were exposed to BDE-153 at 0.9 or 9 mg/kg-day on PND 10. The rat study (Zhang et al. 2013) reported no effect of BDE-153 at 10 mg/kg-day in performance in either the Morris water maze or the passive avoidance test when evaluated at PND 40 or PND 70.

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including lack of randomization of treatment; reduced confidence in outcome assessment due to lack of blinding of outcome assessors; and definitely high risk of bias ratings for not controlling for litter effects in the study design or analysis (see Figure E4-5).

- **Unexplained inconsistencies:** A qualitative evaluation of the evidence suggested a possible downgrade because of the heterogeneity in the evidence. Nevertheless, a meta-analysis of studies of several PBDEs (see Chapter 4 and Appendix E, Section E-5), including BDE-153, and of latency in the last trial of the Morris water maze showed consistent evidence of an effect on this measure of learning, so confidence was not downgraded.
- **Indirectness:** No downgrade because the tests used are considered direct measures of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because no evidence of consistency across species.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

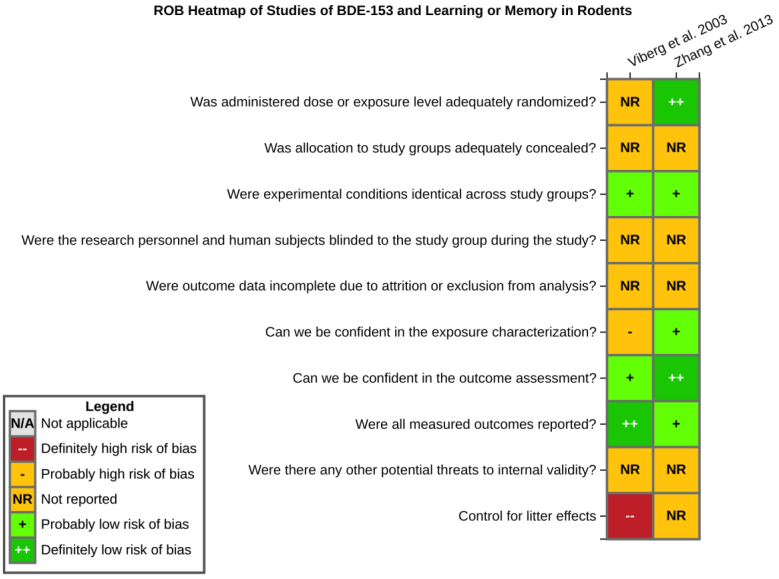


FIGURE E4-5 Risk of bias heatmap of studies of BDE-153 and learning or memory in rodents. In HAWC: https://hawcproject.org/summary/visual/357/.

TABLE E4-6 Studies of BDE-153 and Memory in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Viberg et al. 2003	NMRI mouse	PND 10	PND 180	Morris water maze	0.45	0.9
Zhang et al. 2013	Sprague-Dawley rats	PND 10	PND 40	Morris water maze	1	5
	Sprague-Dawley rats	PND 10	PND 70	Morris water maze	1	5
	Sprague-Dawley rats	PND 10	PND 40	Passive avoidance	10	None
	Sprague-Dawley rats	PND 10	PND 70	Passive avoidance	5	10

NOTE: LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Memory: There is low confidence in the body of evidence on developmental exposure to BDE-153 and effect on memory in rodents. One study in mice and one in rats reported effects. (The two studies are the same ones that tested the effects of BDE-153 on learning.) The mouse study (Viberg et al. 2003) reported longer latencies in the relearning period of mice tested at 6 months of age in the Morris water maze after exposure to BDE-153 at 0.9 or 9 mg/kg-day. The rat study (Zhang et al. 2013) reported increased swimming time in the Morris water maze 1 month after treatment with BDE-153 at 5 and 10 mg/kg-day evaluated at PND 40 or PND 70 and memory impairment in the passive avoidance test at 10 mg/kg-day at PND 70. The two studies report results that were consistent in direction (both decreased performance on a memory test) across species.

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including lack of randomization of treatment; reduced confidence in outcome assessment due to lack of blinding of outcome assessors; and definitely high risk of bias ratings for not controlling for litter effects in the study design or analysis. (See heatmap in Figure E4-5.)
- **Unexplained inconsistencies:** Confidence is usually downgraded if only one study in each tested species is available because the database is insufficient to establish or evaluate consistency for a particular species. Consistency across species (see below), however, would be a reason to upgrade confidence. Considering both factors, there was no downgrade for unexplained inconsistency.
- **Indirectness:** No upgrade because tests were considered direct measures of memory.
- **Imprecision:** No upgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No change, no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** Confidence is usually upgraded if consistent results are observed across species. As noted above, however, it was not possible to evaluate consistency for each species because of the small data set. Considering both factors, there was no upgrade for consistency.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

BDE-203

One study of BDE-203 and effects on learning (see Table E4-7) or memory (see Table E4-8) was found.

TABLE E4-7 Studies of BDE-203 and Learning in Mice

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Viberg et al. 2006	NMRI mouse	PND 3	PND 90	Morris water maze	16.8	None
	NMRI mouse	PND 10	PND 90	Morris water maze	None	16.8

NOTE: LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Learning: There is very low confidence in the data to evaluate whether developmental exposure to BDE-203 affects learning in mice from the single study available. The results suggest that timing of exposure might have an influence on effects because exposure to BDE-203 at 16.8 mg/kg-day on PND 10 affected learning during the acquisition period, whereas exposure on PND 3 had no effect on learning (Viberg et al. 2006).

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including lack of randomization of treatment; reduced confidence in outcome assessment due to lack of blinding of outcome assessors and a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis (see Figure E4-6).
- **Unexplained inconsistencies:** Downgraded because unable to establish or evaluate consistency because the study did not have elements that would strengthen conclusions from a single study, such as multiple species, strains, or particularly large sample sizes (n = 50-100).
- **Indirectness:** No downgrade because test used is considered a direct measure of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because only one dose was tested.
- **Cross-species consistency:** No upgrade because only mice were tested.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

Memory: There is very low confidence in the data to evaluate whether developmental exposure to BDE-203 affects memory in mice because the single study found no effect at a single dose (16.8 mg/kg-day). The study is the same one that also assessed learning.

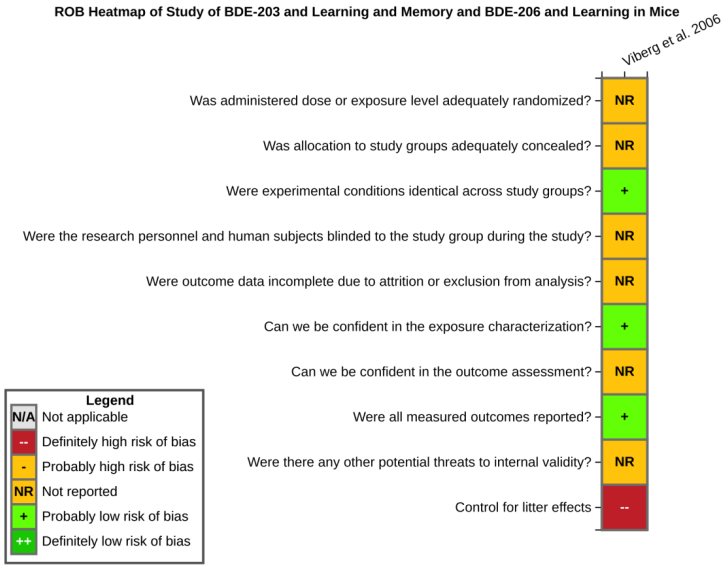


FIGURE E4-6 Risk of bias heatmap of study of BDE-203 and learning and memory and BDE-206 and learning in mice. In HAWC: https://hawcproject.org/summary/visual/358/.

TABLE E4-8 Studies of BDE-203 and Memory in Mice

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Viberg et al. 2006	NMRI mouse	PND 3	PND 90	Morris water maze	16.8	None
	NMRI mouse	PND 10	PND 90	Morris water maze	16.8	None

NOTE: LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including lack of randomization of treatment; reduced confidence in outcome assessment due to lack of blinding of outcome assessors; and a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis. (See heatmap in Figure E4-6.)
- **Unexplained inconsistencies:** Downgraded because unable to establish or evaluate consistency because the study did not have elements that would strengthen conclusions from a single study, such as multiple species, strains, or particularly large sample sizes (n = 50-100).
- **Indirectness:** No downgrade because test used is considered a direct measure of memory.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because only one dose was tested.
- **Cross-species consistency:** No upgrade because only mice were tested.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

BDE-206

One study of BDE-206 and effects on learning was found (see Table E4-9).

Learning: There is very low confidence in the data to evaluate whether developmental exposure to BDE-206 affects learning in mice as the single study found no effect at a single exposure level (16.8 mg/kg-day). The study is the same one that also tested BDE-203 (see above).

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including lack of randomization of treatment; reduced confidence in outcome assessment due to lack of blinding of outcome assessors; and a definitely high risk of bias rating for not controlling for litter effects in the study design or analysis. (see heatmap in Figure E4-6.)
- **Unexplained inconsistencies:** Downgraded because unable to establish or evaluate consistency because the study did not have elements that would strengthen conclusions from a single study, such as multiple species, strains, or particularly large sample sizes (n = 50-100).
- **Indirectness:** No downgrade because test used is considered a direct measure of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because only one dose was tested.
- **Cross-species consistency:** No upgrade because only mice were tested.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

TABLE E4-9 Studies of BDE-206 and Learning in Mice

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Viberg et al. 2006	NMRI mouse	PND 10	PND 90	Morris water maze	16.8	None

NOTE: LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

BDE-209

Studies of BDE-209 and effects on learning (see Table E4-10) and memory (see Table E4-11) were available.

Learning: There is moderate confidence in the body of evidence to evaluate whether developmental exposure to BDE-209 affects learning in rodents. Multiple studies show effects on learning at doses of 20 mg/kg-day or greater when learning was assessed using a Morris water maze; other studies, however, show no effects in the same dose range using other test methods.

- **Risk of bias:** Downgraded because of serious concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors in all studies and a definitely high risk of bias rating in three studies because of failure to control for litter effects in the study design or analysis (see Figure E4-7).

TABLE E4-10 Studies of BDE-209 and Learning in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Biesemeier et al. 2011	Sprague-Dawley rats	GD 6 - PND 21	PND 22, 62	Water T maze	1000	None
Buratovic et al. 2014	NMRI mice	PND 3	5 months	Morris water maze	7.9	None
	NMRI mice	PND 3	7 months	Morris water maze	7.9	None
Chen et al. 2014	Sprague-Dawley rats	GD 1 - 14	PND 25	Morris water maze	10	30
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	10.5	None
Reverte et al. 2013	apoE2 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE2 male mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 male mouse	PND 10	PND 120	Morris water maze	10	30
	apoE4 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 male mouse	PND 10	PND 120	Morris water maze	10	30
	apoE2 female mouse	PND 10	PND 360	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 360	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 360	Morris water maze	30	None
	apoE2 mouse	PND 10	PND 150-180	Cued fear	10	30
	apoE4 mouse	PND 10	PND 150-180	Cued fear	None	10
	apoE4 mouse	PND 10	PND 150-180	Cued fear	30	None
Rice et al. 2009	C57BL6/J mouse	PND 2-15	PND 87	FR performance	20	None
	C57BL6/J mouse	PND 2-15	PND 87	FI performance	20	None
	C57BL6/J mouse	PND 2-15	PND 87	Visual discrimination	6	20
	C57BL6/J mouse	PND 2-15	PND 497	FR performance	None	6
	C57BL6/J mouse	PND 2-15	PND 497	FI performance	20	None
	C57BL6/J mouse	PND 2-15	PND 497	Visual discrimination	None	6
Verma et al. 2013	Swiss albino mouse	PND 3-10	PND 60-66	Morris water maze	None	20
Verma et al. 2014	Swiss albino mouse	PND 3-10	NR	Morris water maze	20	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; NR, not reported; PND, postnatal day.

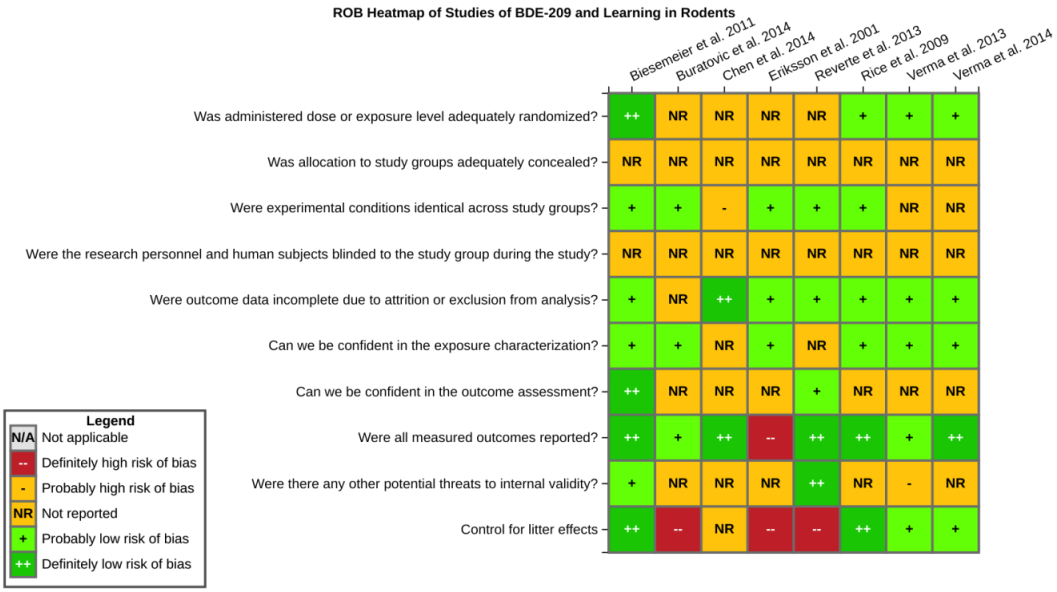


FIGURE E4-7 Risk of bias heatmap of studies of BDE-209 and learning in rodents. In HAWC: https://hawcproject.org/summary/visual/349/.

- **Unexplained inconsistencies:** A qualitative evaluation of the evidence suggested a possible downgrade because of the heterogeneity in the evidence. Nevertheless, a meta-analysis of studies of several PBDEs (see Chapter 4 and Appendix E, Section E-5), including BDE-209, and of latency in the last trial of the Morris water maze showed consistent evidence of an effect on this measure of learning, so confidence was not downgraded.
- **Indirectness:** No downgrade because tests used were considered direct measures of learning.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because of inconsistencies in the results between rat and mouse studies.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

Memory: There is low confidence in the body of evidence to evaluate whether developmental exposure to BDE-209 affects learning in rodents. Multiple mouse studies show effects on memory at doses of 3.4 mg/kg-day or greater when memory was assessed using a Morris water maze; other studies, however, show no effects on memory at the same dose range using other methods.

- **Risk of bias:** Downgraded because of serious concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors in all studies and a definitely high risk of bias rating in three studies because of failure to control for litter effects in the study design or analysis (see Figure E4-8).
- **Unexplained inconsistencies:** Downgraded for inconsistency.
- **Indirectness:** No downgrade because tests used were considered direct measures of memory.

TABLE E4-11 Studies of BDE-209 and Memory in Rodents

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Biesemeier et al. 2011	Sprague-Dawley rats	GD6-PND 21	PND 22, 62	Water T maze	1000	None
Buratovic et al. 2014	NMRI mice	PND 3	5 months	Morris water maze	None	3.4
	NMRI mice	PND 3	7 months	Morris water maze	None	3.4
Eriksson et al. 2001	NMRI mice	PND 10	5 months	Morris water maze	None	10.5
Reverte et al. 2013	apoE2 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE2 male mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 male mouse	PND 10	PND 120	Morris water maze	None	10
	apoE4 female mouse	PND 10	PND 120	Morris water maze	30	None
	apoE4 male mouse	PND 10	PND 120	Morris water maze	None	10
	apoE2 female mouse	PND 10	PND 360	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 360	Morris water maze	30	None
	apoE4 female mouse	PND 10	PND 360	Morris water maze	30	None
Verma et al. 2013	Swiss albino mouse	PND 3-10	PND 60-66	Morris water maze	None	20
	Swiss albino mouse	PND 3-10	PND 60-66	Radial maze	None	20
	Swiss albino mouse	PND 3-10	PND 60-66	Radial maze	None	20
Verma et al. 2014	Swiss albino mouse	PND 3-10	NR	Morris water maze	20	None
	Swiss albino mouse	PND 3-10	NR	Radial maze	20	None
	Swiss albino mouse	PND 3-10	NR	Radial maze	20	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; NR, not reported; PND, postnatal day.

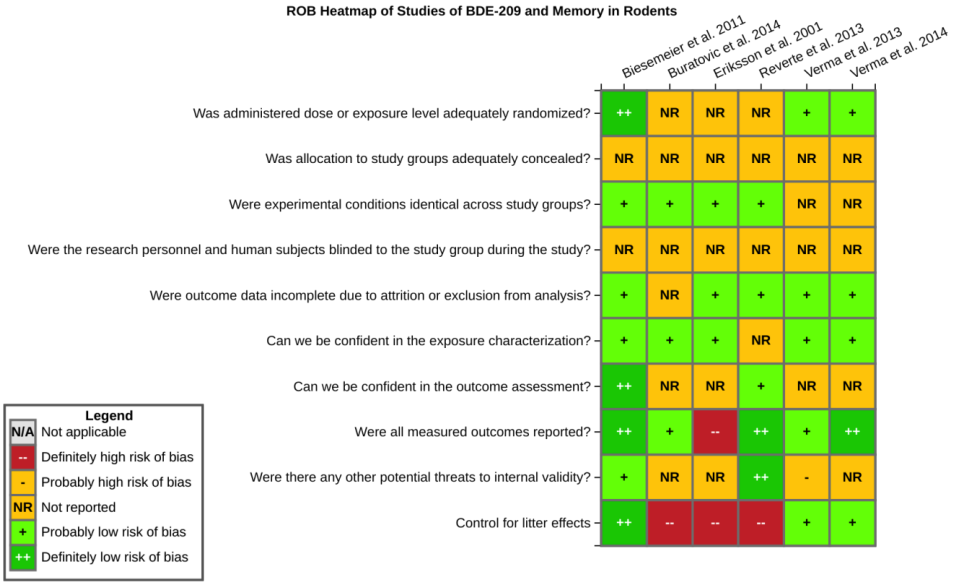


FIGURE E4-8 Risk of bias heatmap of studies of BDE-209 and memory in rodents. In HAWC: https://hawcproject.org/summary/visual/350/.

- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because of inconsistencies in the results between rat and mouse studies.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

DE-71

Studies of DE-71 and effects on learning (see Table E4-12), memory (see Table E4-13), and attention (see Table E4-14) were available.

Learning: There is very low confidence in the body of evidence on developmental exposure to DE-71 and effects on learning in rats. The results of the three available studies were inconsistent and used different tests (Morris water maze, radial maze, and visual discrimination) and animals of different ages. One study (Dufault et al. 2005) reported increased errors in the visual discrimination task at the single dose tested (30 mg/kg-day). The other two studies reported no effects of DE-71 on learning at the same dose using longer exposure windows; however, the animals in these studies were evaluated at older ages than were the rats tested in the Dufault et al. (2005) study.

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors and a definitely high risk of bias rating for exposure characterization in two of the studies (see Figure E4-9).

TABLE E4-12 Studies of DE-71 and Learning in Rats

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Bowers et al. 2015	Sprague-Dawley rats	GD 1 - PND 21	PND 235	Morris water maze	30	None
de-Miranda et al. 2016	Wistar rats	PND 5-22	PND 100	Radial maze	30	None
Dufault et al. 2005	Long-Evans rats	PND 6-12	PND 30	Visual discrimination	None	30

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

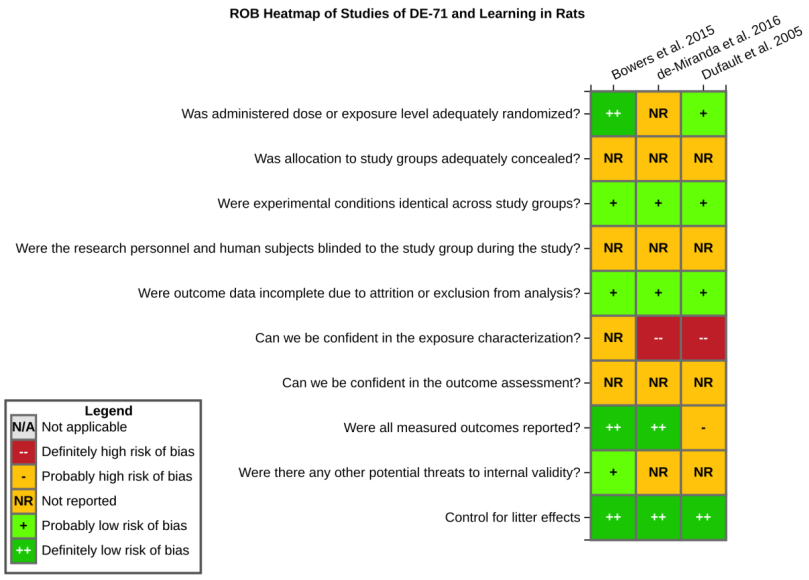


FIGURE E4-9 Risk of bias heatmap of studies of DE-71 and learning in rats. In HAWC: https://hawcproject.org/summary/visual/344/.

TABLE E4-13 Studies of DE-71 and Memory in Rats

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Bowers et al. 2015	Sprague-Dawley rats	GD 1 - PND 21	PND 235	Morris water maze	30	None
de-Miranda et al. 2016	Wistar rats (female)	PND 5-22	PND 100	Radial maze	None	30
	Wistar rats (male)	PND 5-22	PND 100	Radial maze	30	None

NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Memory: There is very low confidence in the body of evidence on developmental exposure to DE-71 and effects on memory in rats. The results of the two available studies were inconsistent and were evaluated in animals of different ages and with different tests (Morris water maze and radial maze). One study (de-Miranda et al. 2016) reported a reference memory deficit in female Wistar rats (not males) in the radial maze at the single dose tested (30 mg/kg-day).

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors and a definitely high risk of bias rating for exposure characterization in one of the studies (see Figure E4-10).
- **Unexplained inconsistencies:** Downgrade for inconsistency.
- **Indirectness:** No downgrade because tests used are considered direct measures of memory.
- **Imprecision:** No downgrade because no or minimal indications of large standard deviations (i.e., SD > mean).
- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because only rats were tested.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

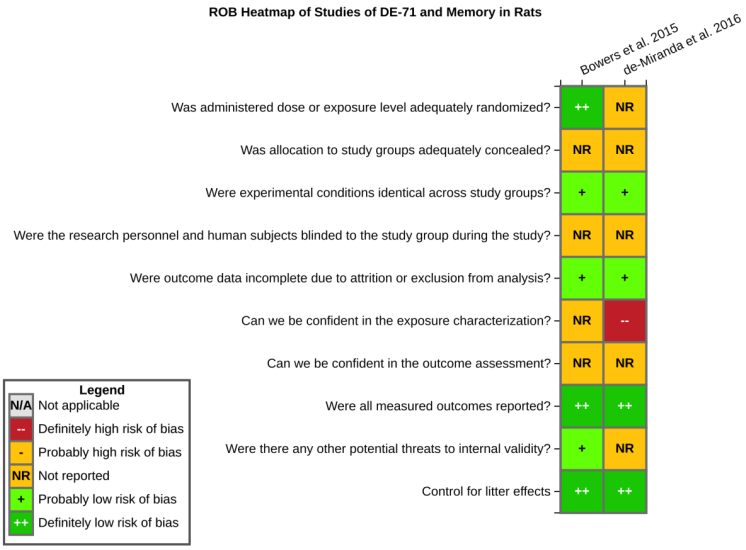


FIGURE E4-10 Risk of bias heatmap of studies of DE-71 and memory in rats. In HAWC: https://hawcproject.org/summary/visual/345/.

TABLE E4-14 Studies of DE-71 and Attention in Rats

Study	Species	Life Stage Exposed	Observation Time	Test	NOAEL (mg/kg-day)	LOAEL (mg/kg-day)
Driscoll et al. 2009	Long-Evans rats	Lifetime*	PND 40-95	Visual task	4.5	None
	Long-Evans rats	Lifetime*	PND 40-95	Attention task I	3	4.5
	Long-Evans rats	Lifetime*	PND 40-95	Attention task II	4.5	None
Driscoll et al. 2012	Long-Evans rats	PND 6-12	PND 40-95	Attention task	15	None
	Long-Evans rats	PND 6-12	PND 40-95	Visual task	15	None
Dufault et al. 2005	Long-Evans rats	PND 6-12	PND 30	Attention task	30	None

*Animals were exposed throughout gestation via treated dams and via chow after weaning.
NOTE: GD, gestation day; LOAEL, lowest-observed-adverse-effect level; NOAEL, no-observed-adverse-effect level; PND, postnatal day.

Attention: There is very low confidence in the body of evidence on developmental exposure to DE-71 and effects on attention in rats. All of the data are from a single laboratory (Dufault et al. 2005; Driscoll et al. 2009, 2012) and the majority of the tests reported no effects at doses up to 30 mg/kg-day across multiple tests (various attention tasks and a visual task). In one experiment (Driscoll et al. 2009), rats exposed to DE-71 at 4.5 mg/kg-day demonstrated lower accuracy in Attention Task 1.

- **Risk of bias:** Downgraded twice because of serious concerns about multiple risk of bias issues, including reduced confidence in outcome assessment due to lack of blinding of outcome assessors and a definitely high risk of bias rating for exposure characterization in one of the studies (see Figure E4-11).
- **Unexplained inconsistencies:** Downgrade for inconsistency.
- **Indirectness:** No downgrade because tests used are considered direct measures of attention.
- **Imprecision:** No downgraded because no or minimal indications of large standard deviations (i.e., SD > mean).

- **Dose-response:** No upgrade because no clear evidence of dose-response gradient within or across studies.
- **Cross-species consistency:** No upgrade because only rats were tested.
- **Other potential downgrades or upgrades:** No evidence of publication bias (see Section E-3), large magnitude of effect, or residual confounding or other related factors that would affect confidence in the estimated effect.

FIGURE E4-11 Risk of bias heatmap of studies of DE-71 and attention in rats. In HAWC: https://hawcproject.org/summary/visual/347/.

SECTION E-5

Supporting Information for the Meta-Analyses of Studies of PBDEs

Meta-Analyses on Combined Data on PBDEs

TABLE E5-1 Overall Analyses and Sensitivity Analyses of Studies of PBDEs and Latency in Last Trial of the Morris Water Maze

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	25.76	20.32	31.19	0.000000	4.65	24.46	0.2685	101.48
Trend in log10(dose)	log10(dose)	5.74	-2.16	13.63	0.154334	3.41	14.58	0.3825	96.49*
Linear in dose10	dose10	9.61	3.79	15.42	0.001209	17.12	81.59	0.0000	116.50
Linear-Quadratic in dose10	dose10	28.07	11.22	44.91	0.001093	13.92	72.33	0.0001	108.36
	I(dose10^2)	-4.45	-8.30	-0.60	0.023373				
Sensitivity Analyses									
Overall minus Viberg et al. 2003	intrcpt	25.77	20.07	31.46	0.000000	4.95	32.02	0.1180	77.61
Overall minus Chen et al. 2014	intrcpt	25.60	18.33	32.86	0.000000	5.17	22.40	0.3185	81.54
Overall minus Verma et al. 2013	intrcpt	25.52	20.08	30.97	0.000000	4.65	26.04	0.3070	90.62
Overall minus He et al. 2011	intrcpt	27.13	20.03	34.22	0.000000	4.14	13.54	0.4338	81.26
Overall minus Woods et al. 2012	intrcpt	25.87	20.38	31.36	0.000000	4.64	27.53	0.2743	82.77
Overall minus Cheng et al. 2009	intrcpt	25.12	19.24	30.99	0.000000	5.03	27.26	0.2377	94.83
Highest Doses-Overall	intrcpt	32.95	26.67	39.23	0.000000	0.00	0.00	0.6596	56.49
Highest Doses-Overall minus Viberg et al. 2003	intrcpt	32.91	26.55	39.26	0.000000	0.00	0.00	0.5325	50.67
Highest Doses-Overall minus Chen et al. 2014	intrcpt	32.39	24.92	39.86	0.000000	0.00	0.00	0.5419	52.83
Highest Doses-Overall minus Verma et al. 2013	intrcpt	32.67	26.38	38.97	0.000000	0.00	0.00	0.7322	47.94
Highest Doses-Overall minus He et al. 2011	intrcpt	33.34	24.55	42.14	0.000000	0.00	0.00	0.5336	53.08
Highest Doses-Overall minus Woods et al. 2012	intrcpt	33.32	26.95	39.70	0.000000	0.00	0.00	0.8448	45.18
Highest Doses-Overall minus Cheng et al. 2009	intrcpt	33.13	26.28	39.98	0.000000	0.00	0.00	0.5337	52.46
Highest Doses-Trend in log10(dose)	log10(dose)	3.40	-6.47	13.28	0.499553	0.00	0.00	0.5978	70.61
Highest Doses-Linear in dose10	dose10	10.01	0.06	19.95	0.048547	26.04	88.45	0.0000	67.93
Highest Doses-Linear-Quadratic in dose10	dose10	48.56	17.03	80.08	0.002536	15.99	55.56	0.0236	76.79
	I(dose10^2)	-8.34	-14.87	-1.81	0.012356				

*Indicates the lowest AICc.

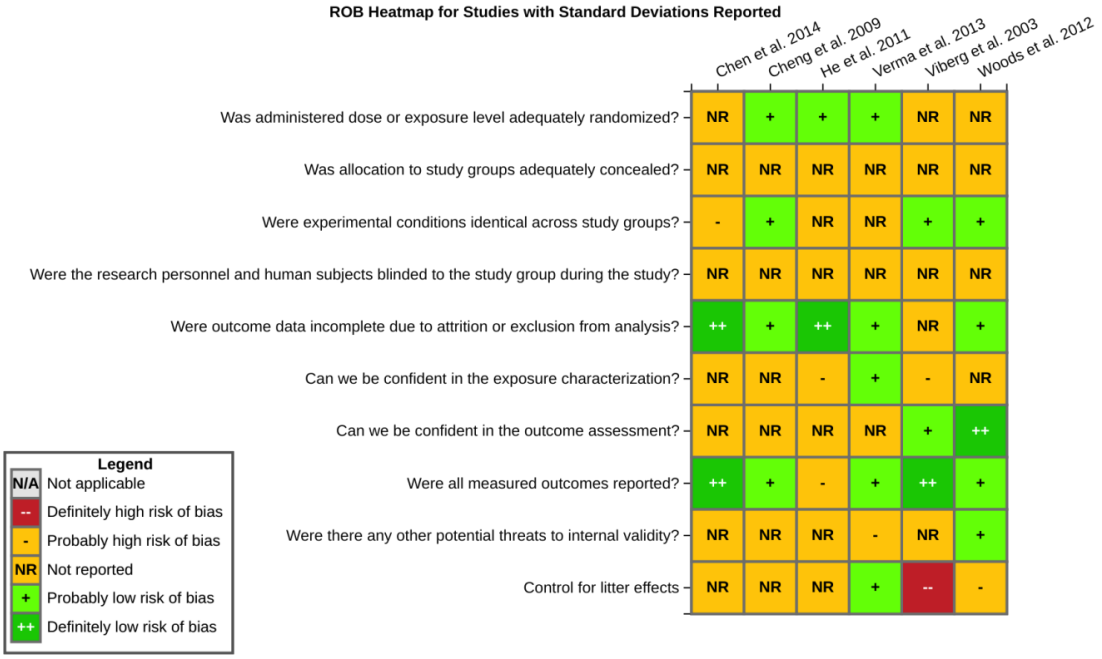


FIGURE E5-1 Risk of bias heatmap of studies of PBDEs and latency in last trial of the Morris water maze with standard deviations reported or digitized from figures in the publication. In HAWC: https://hawcproject.org/summary/visual/364/.

FIGURE E5-2 Risk of bias heatmap of studies of PBDEs and latency in last trial of the Morris water maze without standard deviations. In HAWC: https://hawcproject.org/summary/visual/365/.

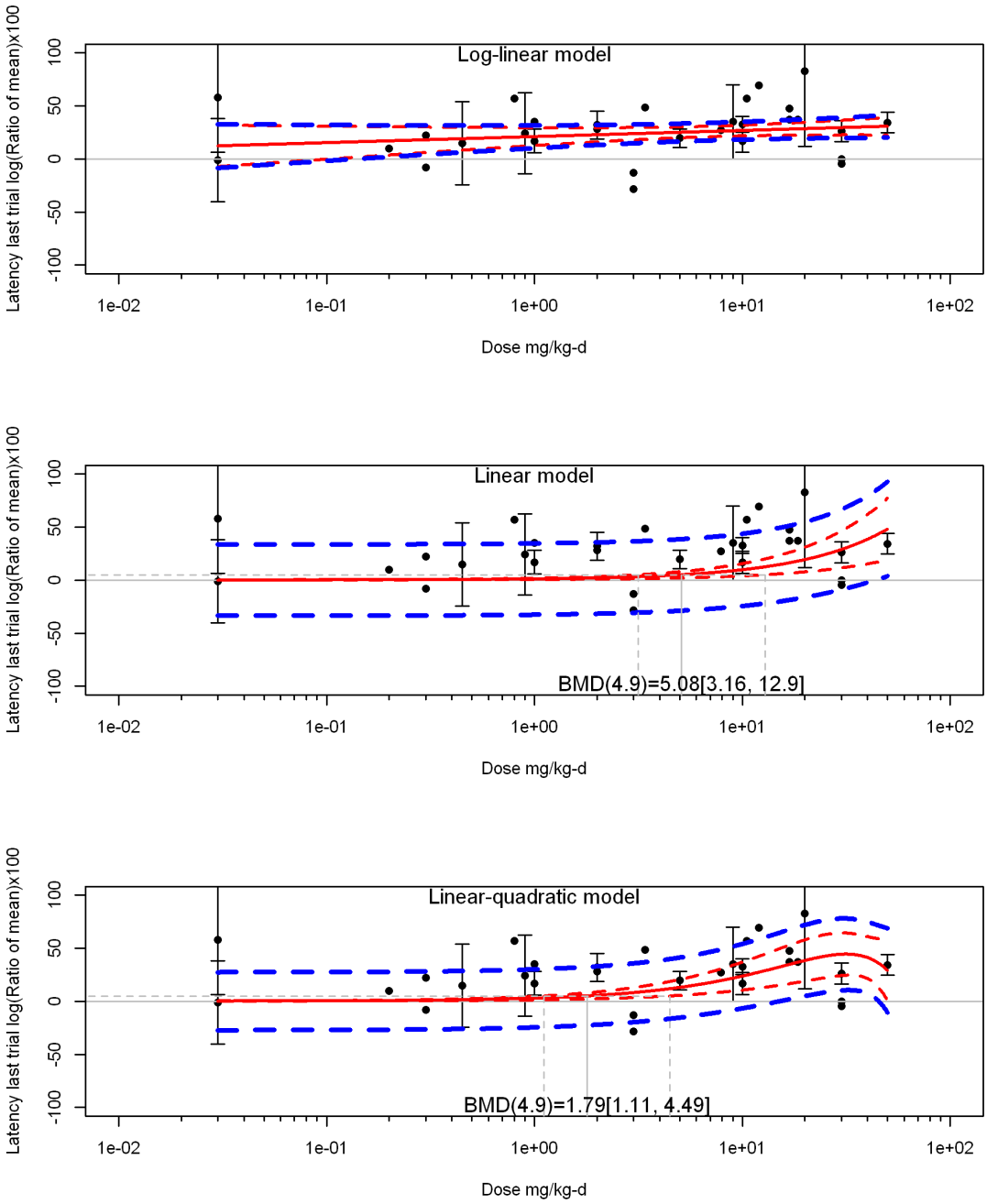


FIGURE E5-3 Benchmark dose estimates from studies of PBDEs and latency in last trial of the Morris water maze in rats and mice. Points without error bars are studies for which a standard deviation was not reported or could not be digitized from figures in the publication. They were not included in the model fitting but are shown for comparison. There is no obvious bias between studies that reported standard deviations and those that did not.

Meta-Analyses on Individual PBDEs

BDE-47

- Statistically significant overall effect. Heterogeneity (I^2 = 44%), but it was not statistically significant. Overall effect was robust to using only highest dose from each study.

- Positive trends in $\log_{10}$(dose) and in dose, but only the latter was statistically significant. Reduced heterogeneity for $\log_{10}$(dose) and linear model. Benchmark dose for a 5% change was estimated to be 1.4 mg/kg-day (95% CI: 1.0, 2.4) from the linear model and 0.83 mg/kg-day (95% CI: 0.34, 5.9) from the linear-quadratic model.

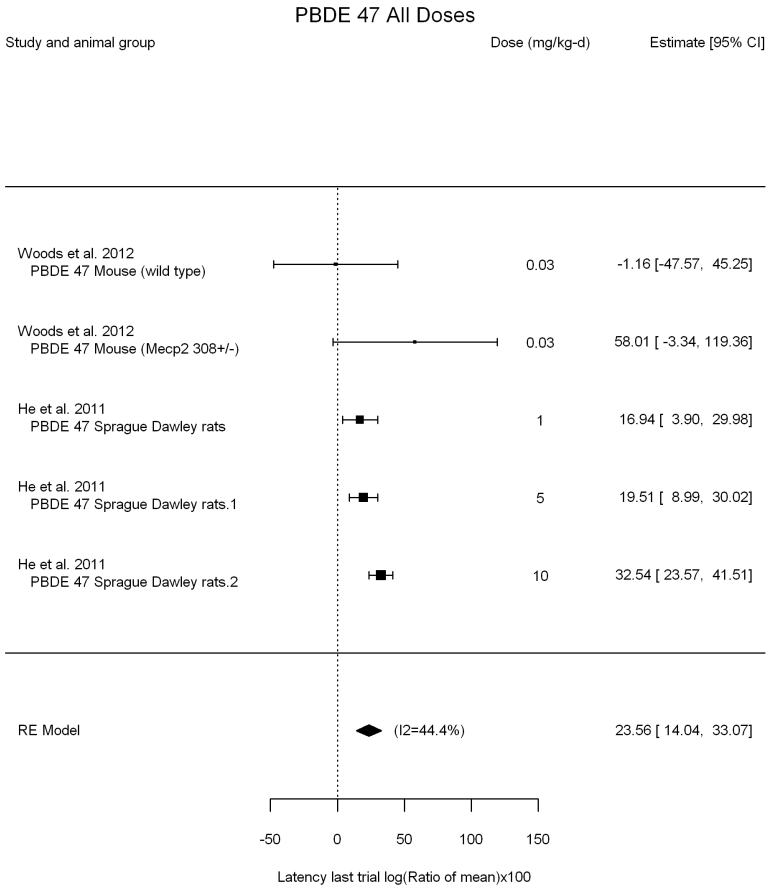


FIGURE E5-4 Results of meta-analysis of studies of BDE-47 and latency in last trial of the Morris water maze.

TABLE E5-2 Overall Analyses and Sensitivity Analyses of Studies BDE-47 and Latency in Last Trial of the Morris Water Maze

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	23.56	14.04	33.07	0.0000	6.72	44.43	0.1092	49.01*
Trend in log10(dose)	log10(dose)	7.19	-6.23	20.61	0.2939	4.72	27.79	0.1610	54.64
Linear in dose10	dose10	34.55	20.13	48.97	0.0000	6.76	41.56	0.0957	50.90
Linear-Quadratic in dose10	dose10	60.96	-27.12	149.04	0.1750	10.24	48.02	0.0829	58.02
Linear-Quadratic in dose10	I(dose10^2)	-29.19	-124.65	66.27	0.5490	10.24	48.02	0.0829	58.02
Sensitivity Analyses									
Highest Doses-Overall	intrcpt	31.87	23.15	40.59	0.0000	0.00	0.00	0.2638	34.52

*Indicates the lowest AICc.

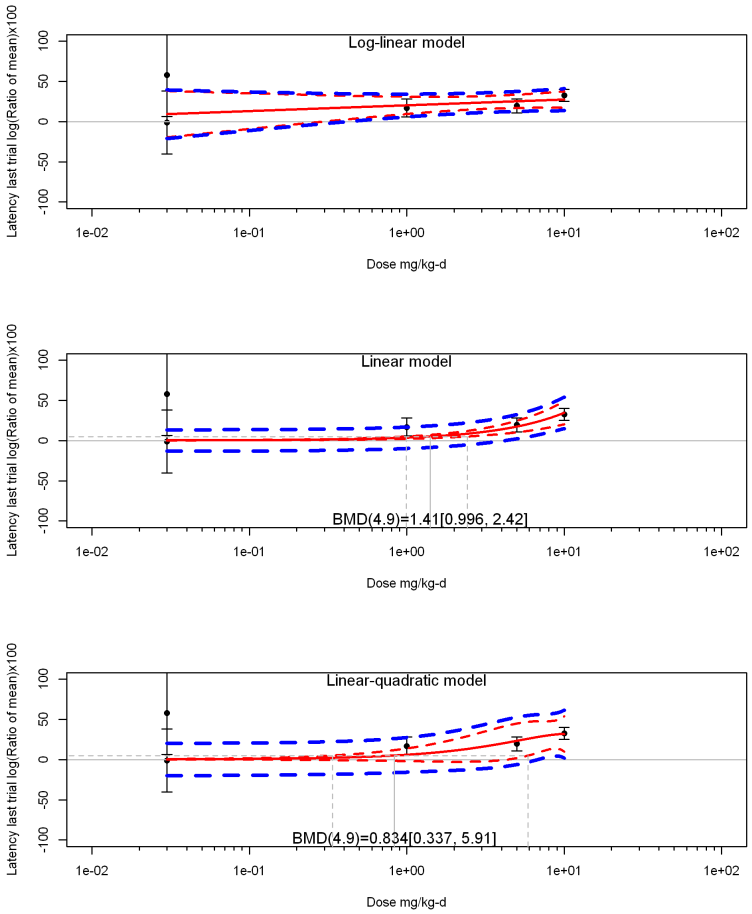


FIGURE E5-5 Benchmark dose estimates from studies of BDE-47 and latency in last trial of the Morris water maze.

BDE-153

- Statistically significant overall effect; no heterogeneity. Too few data for a sensitivity analysis.
- Positive trend, but not a statistically significant trend. Only central estimate and lower bound could be estimated for a benchmark dose for a 5% change: 1.2 mg/kg-day (95% CI: 0.6, >10).

TABLE E5-3 Overall Analyses and Sensitivity Analyses of Studies BDE-153 and Latency in Last Trial of the Morris Water Maze

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	25.40	-0.18	50.99	0.052	0	0	0.82	32.56*
Trend in log10(dose)	log10(dose)	14.03	-30.94	58.99	0.541	0	0	0.88	38.16
Linear in dose10	dose10	41.17	-4.69	87.04	0.078	0	0	0.58	33.38
Linear-Quadratic in dose10	dose10	304.47	-193.63	802.57	0.231	0	0	0.93	38.17
Linear-Quadratic in dose10	I(dose10^2)	-295.30	-851.58	260.97	0.298	0	0	0.93	38.17

*Indicates the lowest AICc.

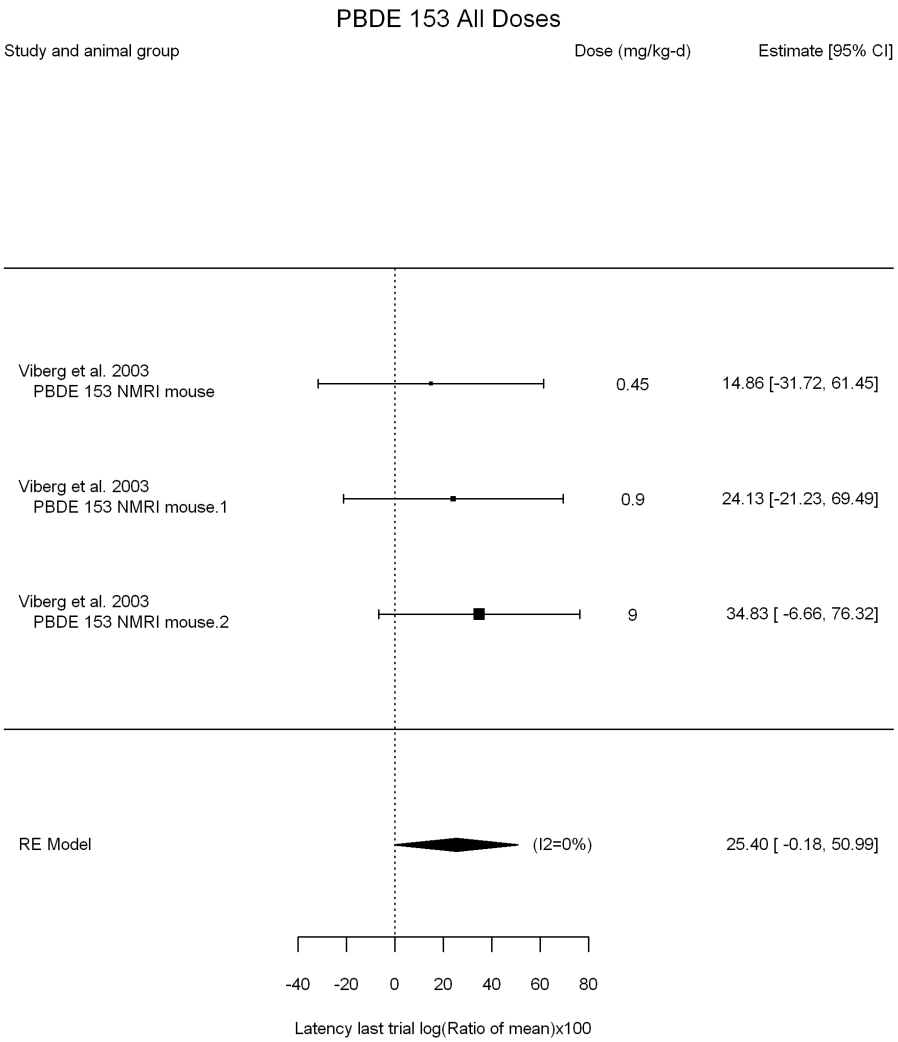


FIGURE E5-6 Results of meta-analysis of studies of BDE-153 and latency in last trial of the Morris water maze.

BDE-209

- Statistically significant overall effect. Heterogeneity ($I^2 = 42\%$), but it was not statistically significant. Overall effect was robust to using only highest dose from each study (only two studies).
- Statistically significant trends in $\log_{10}$(dose) and linear trend in dose, with reduced heterogeneity. BMD estimates for a 5% change were 6.3 mg/kg-day (95% CI: 4.8, 9.2) from the linear model and 3.5 mg/kg-day (95% CI: 2.2, 7.9) from the linear-quadratic model.

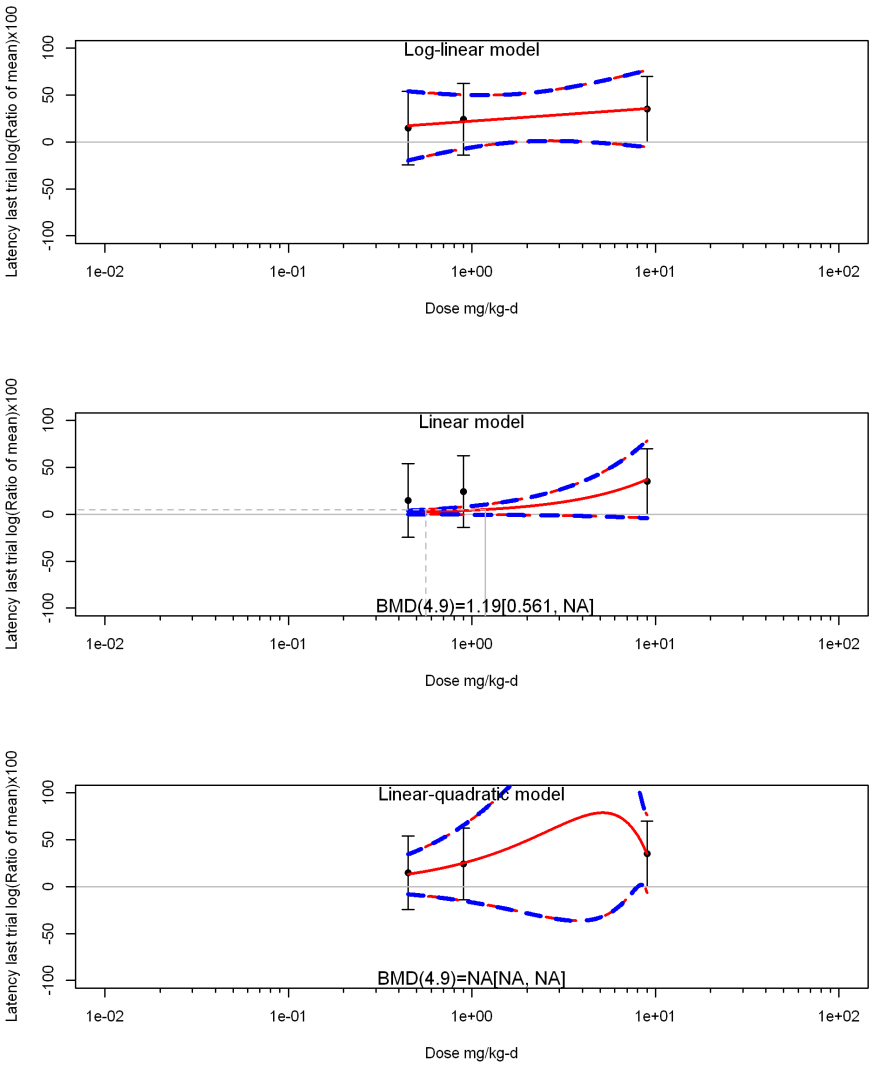


FIGURE E5-7 Benchmark dose estimates from studies of BDE-153 and latency in last trial of the Morris water maze in rats and mice.

TABLE E5-4 Overall Analyses and Sensitivity Analyses of Studies BDE-209 and Latency in Last Trial of the Morris Water Maze

Analysis	Estimate	Beta	CI, Lower Bound	CI, Upper Bound	P value	tau	I^2	P value for Heterogeneity	AICc
Primary Analyses									
Overall	intrcpt	26.69	16.79	36.60	0.00000	6.36	42.27	0.12	41.15*
Trend in log10(dose)	log10(dose)	23.92	0.01	47.83	0.04990	0.00	0.00	0.37	46.48
Linear in dose10	dose10	7.70	5.32	10.08	0.00000	4.01	22.14	0.16	41.33
Linear-Quadratic in dose10	dose10	14.68	5.98	23.39	0.00094	0.00	0.00	0.27	46.94
Linear-Quadratic in dose10	I(dose10^2)	-1.60	-3.53	0.33	0.10377	0.00	0.00	0.27	46.94
Sensitivity Analyses									
Highest Doses-Overall	intrcpt	39.61	9.96	69.26	0.00883	14.90	18.95	0.27	25.90

*Indicates the lowest AICc.

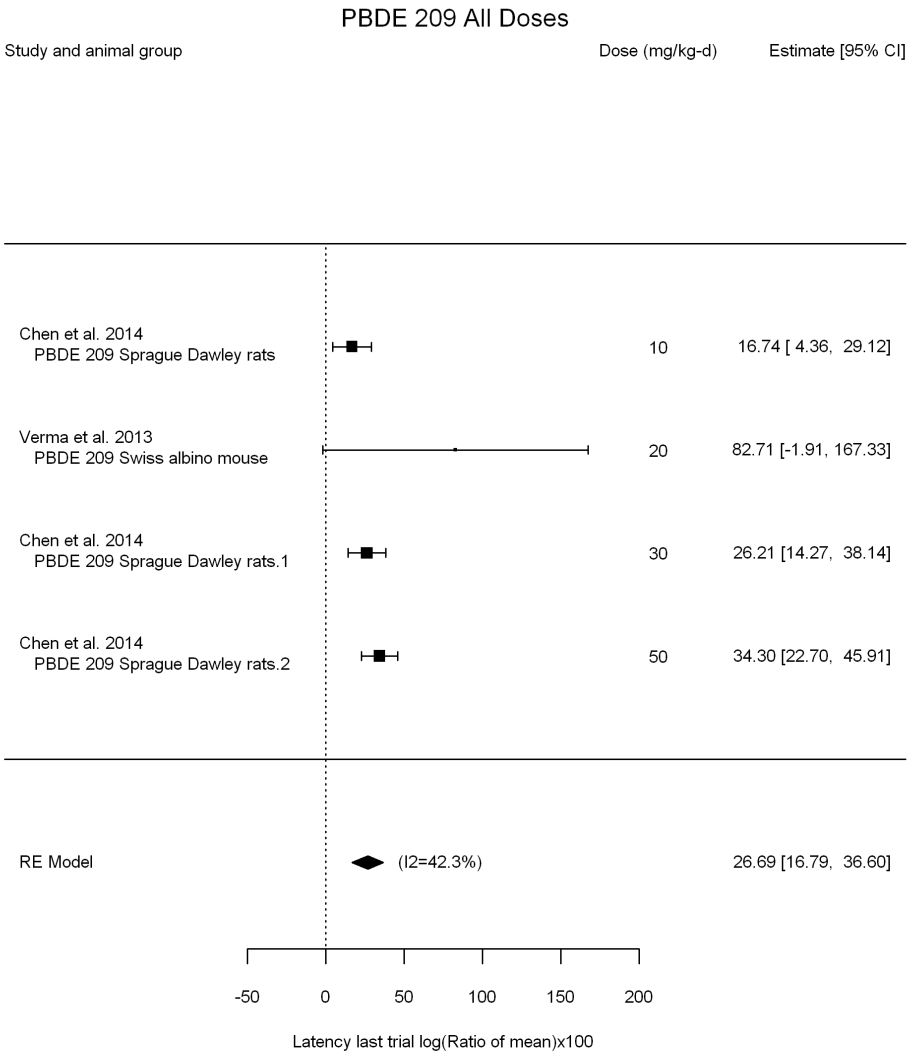


FIGURE E5-8 Results of meta-analysis of studies of BDE-209 and latency in last trial of the Morris water maze.

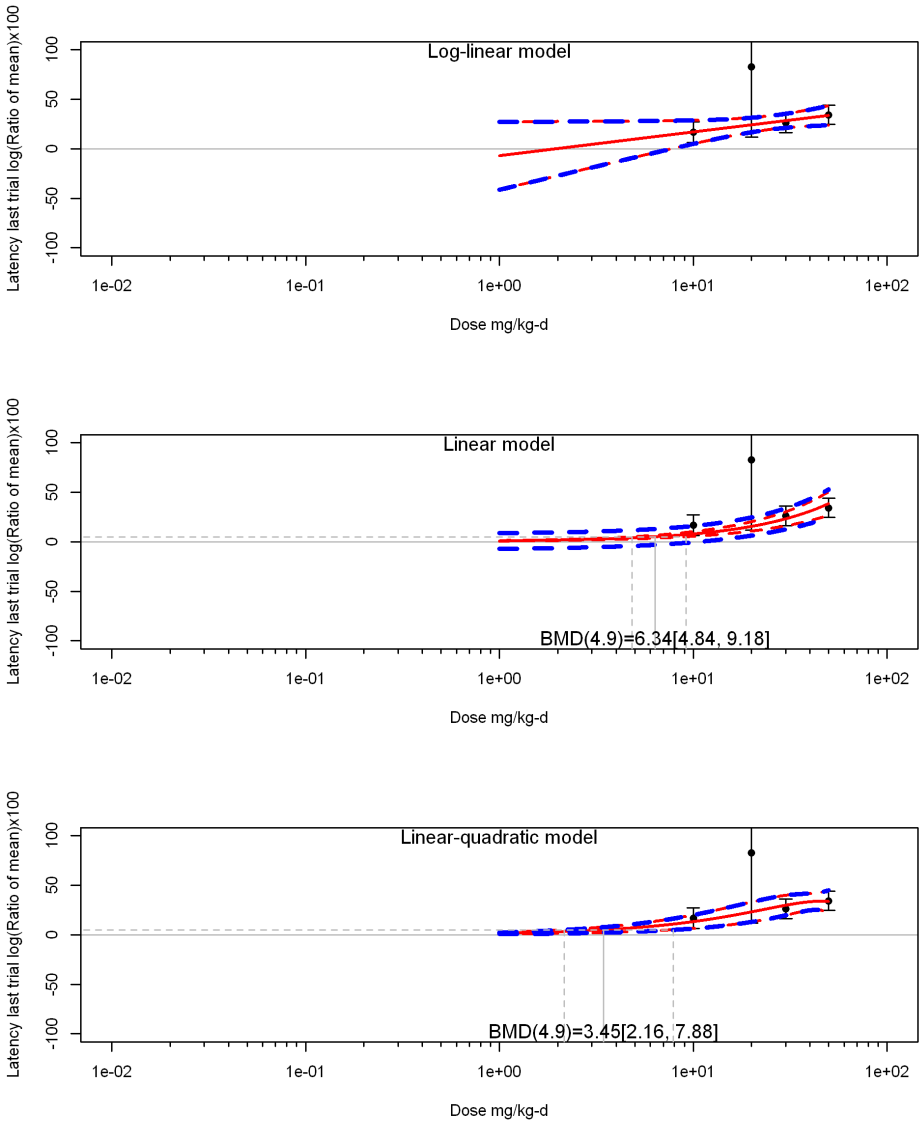


FIGURE E5-9 Benchmark dose estimates from studies of BDE-209 and latency in last trial of the Morris water maze.

REFERENCES

Biesemeier, J.A., M.J. Beck, H. Silberberg, N.R. Myers, J.M. Ariano, A. Radovsky, L. Freshwater, D.W. Sved, S. Jacobi, D.G. Stump, M.L. Hardy, and T. Stedeford. 2011. An oral developmental neurotoxicity study of decabromodiphenyl ether (DecaBDE) in rats. Birth Defects Res. B Dev. Reprod. Toxicol. 92(1):17-35.

Blanco, J., M. Mulero, L. Heredia, A. Pujol, J.L. Domingo, and D.J. Sánchez**.** 2013. Perinatal exposure to BDE-99 causes learning disorders and decreases serum thyroid hormone levels and BDNF gene expression in hippocampus in rat offspring. Toxicology 308:122-128.

Bowers, W.J., P.M. Wall, J.S. Nakai, A. Yagminas, M. Wade, and N. Li. 2015. Behavioral and thyroid effects of in utero and lactational exposure of Sprague-Dawley rats to the polybrominated diphenyl ether mixture DE71. Neurotoxicol. Teratol. 52(Pt. B):127-142.

Buratovic, S., H. Viberg, A. Fredriksson, and P. Eriksson. 2014. Developmental exposure to the polybrominated diphenyl ether PBDE 209: Neurobehavioural and neuroprotein analysis in adult male and female mice. Environ. Toxicol. Pharmacol. 38(2):570-585.

Chen, A., K. Yolton, S.A. Rauch, G.M. Webster, R. Hornung, A. Sjödin, K.N. Dietrich, and B.P. Lanphear. 2014. Prenatal polybrominated diphenyl ether exposures and neurodevelopment in U.S. children through 5 years of age: The HOME study. Environ. Health Perspect. 122(8):856-862.

Cheng, J., J. Gu, J. Ma, X. Chen, M. Zhang, and W. Wang. 2009. Neurobehavioural effects, redox responses and tissue distribution in rat offspring developmental exposure to BDE-99. Chemosphere 75(7):963-968.

de-Miranda, A.S., S.N. Kuriyama, C.S. da-Silva, M.S. do-Nascimento, T.E. Parente, and F.J. Paumgartten. 2016. Thyroid hormone disruption and cognitive impairment in rats exposed to PBDE during postnatal development. Reprod. Toxicol. 63:114-124.

Driscoll, L.L., A.M. Gibson, and A. Hieb. 2009. Chronic postnatal DE-71 exposure: Effects on learning, attention and thyroxine levels. Neurotoxicol. Teratol. 31(2):76-84.

Driscoll, L.L., J. Kaplan, E. Bucuvalas, H. Allen, J. Kraut, and J. Fitzpatrick. 2012. Acute postnatal exposure to the pentaBDE commercial mixture DE-71 at 5 or 15mg/kg/day does not produce learning or attention deficits in rats. Neurotoxicol. Teratol. 34(1):20-26.

Dufault, C., G. Poles, and L.L. Driscoll. 2005. Brief postnatal PBDE exposure alters learning and the cholinergic modulation of attention in rats. Toxicol. Sci. 88(1):172-180.

Eriksson, P., E. Jakobsson, and A. Fredriksson. 2001. Brominated flame retardants: A novel class of developmental neurotoxicants in our environment? Environ. Health Perspect. 109(9):903-908.

Fischer, C., A. Fredriksson, and P. Eriksson. 2008. Coexposure of neonatal mice to a flame retardant PBDE 99 (2, 2',4, 4',5-pentabromodiphenyl ether) and methyl mercury enhances developmental neurotoxic defects. Toxicol. Sci. 101(2):275-285.

He, P., A.G. Wang, T. Xia, P. Gao, Q. Niu, L.J. Guo, and X.M. Chen. 2009. Mechanisms underlying the developmental neurotoxic effect of PBDE-47 and the enhanced toxicity associated with its combination with PCB153 in rats. NeuroToxicology 30(6):1088-1095.

He, P., A. Wang, Q. Niu, L. Guo, T. Xia, and X. Chen. 2011. Toxic effect of PBDE-47 on thyroid development, learning, and memory, and the interaction between PBDE-47 and PCB153 that enhances toxicity in rats. Toxicol. Ind. Health 27(3):279-288.

Koenig, C.M., J. Lango, I.N. Pessah, and R.F. Berman. 2012. Maternal transfer of BDE-47 to offspring and neurobehavioral development in C57BL/6J mice. Neurotoxicol. Teratol. 34(6):571-580.

Llansola, M., M. Hernandez-Viadel, S. Erceg, C. Montoliu, and V. Felipo. 2009. Increasing the function of the glutamate-nitric oxide-cyclic guanosine monophosphate pathway increases the ability to learn a Y-maze task. J. Neurosci. Res. 87(10):2351-2355.

Reverte, I., A.B. Klein, J.L. Domingo, and M.T. Colomina. 2013. Long term effects of murine postnatal exposure to decabromodiphenyl ether (BDE-209) on learning and memory are dependent upon APOE polymorphism and age. Neurotoxicol. Teratol. 40:17-27.

Rice, D. C., W.D. Thompson, E.A. Reeve, K.D. Onos, M. Assadollahzadeh, and V.P. Markowski. 2009. Behavioral changes in aging but not young mice after neonatal exposure to the polybrominated flame retardant decaBDE. Environ. Health Perspect. 117(12):1903-1911.

Ta, T.A., C.M. Koenig, M.S. Golub, I.N. Pessah, L. Qi, P.A. Aronov, and R.F. Berman. 2011. Bioaccumulation and behavioral effects of 2,2',4,4'-tetrabromodiphenyl ether (BDE-47) in perinatally exposed mice. Neurotoxicol. Teratol. 33(3):393-404.

Verma, P., P. Singh, and B.S. Gandhi. 2013. Prophylactic efficacy of *Bacopa monnieri* on decabromodiphenyl ether (PBDE-209)-induced alterations in oxidative status and spatial memory in mice. Asian J. Pharm. Clin. Res. 6(3):242-247.

Verma, P., P. Singh, and B.S. Gandhi. 2014. Neuromodulatory role of *Bacopa monnieri* on oxidative stress induced by postnatal exposure to decabromodiphenyl ether (PBDE -209) in neonate and young female mice. Iran. J. Basic Med. Sci. 17(4):307-311.

Viberg, H., A. Fredriksson, and P. Eriksson. 2003. Neonatal exposure to polybrominated diphenyl ether (PBDE 153) disrupts spontaneous behaviour, impairs learning and memory, and decreases hippocampal cholinergic receptors in adult mice. Toxicol. Appl. Pharmacol. 192(2):95-106.

Viberg, H., N. Johansson, A. Fredriksson, J. Eriksson, G. Marsh, and P. Eriksson. 2006. Neonatal exposure to higher brominated diphenyl ethers, hepta-, octa-, or nonabromodiphenyl ether, impairs spontaneous behavior and learning and memory functions of adult mice. Toxicol. Sci. 92(1):211-218.

Woods, R., R.O. Vallero, M.S. Golub, J.K. Suarez, T.A. Ta, D.H. Yasui, L.H. Chi, P.J. Kostyniak, I.N. Pessah, R.F. Berman, and J.M. LaSalle. 2012. Long-lived epigenetic interactions between perinatal PBDE exposure and Mecp2308 mutation. Hum. Mol. Genet. 21(11):2399-2411.

Zhang, H., X. Li, J. Nie, and Q. Niu. 2013. Lactation exposure to BDE-153 damages learning and memory, disrupts spontaneous behavior and induces hippocampus neuron death in adult rats. Brain Res. 1517:44-56.

Zhao, W., J. Cheng, J. Gu, Y. Liu, M. Fujimura, and W. Wang. 2014. Assessment of neurotoxic effects and brain region distribution in rat offspring prenatally co-exposed to low doses of BDE-99 and methylmercury. Chemosphere 112:170-176.

Appendix F

Supporting Materials for the PBDE (Human) Systematic Review

SECTION F-1

PBDE (HUMAN) PROTOCOL TO UPDATE SYSTEMATIC REVIEW OF THE ASSOCIATION BETWEEN DEVELOPMENTAL EXPOSURES TO PBDES AND HUMAN NEURODEVELOPMENT

August 3, 2016
(Modified on November 11, 2016—See Section F-1c)

BACKGROUND AND INTRODUCTION

Polybrominated biphenyl ethers (PBDEs) are synthetic brominated flame retardants that are ubiquitous environmental contaminants that have been measured in animals and in humans. They have been linked to neurological impairments after developmental exposure in animal and in human studies. During the course of exploring this class of chemicals, the committee learned of other systematic reviews on this topic and decided that using one of them could provide a case study of how to evaluate an existing review for risk of bias and how to update an existing review.

OBJECTIVE AND SPECIFIC AIMS

Review Question

The overall objective of this systematic review is to answer the question is developmental exposure to PBDEs in humans associated with alterations in quantitative measures of intelligence or attention-deficit/hyperactivity disorder (ADHD) and attention-related behavioral conditions?

The specific aims of the review are to:

- Evaluate a systematic review using the risk of bias tool ROBIS (Whiting et al. 2016).
- Update an existing review by doing the following:
  - Identify new literature reporting the effects of developmental exposure to PBDEs on measures of intelligence or on ADHD and attention-related behavioral conditions.
  - Extract data on the effects of developmental exposure to PBDEs on alterations in quantitative measures of intelligence or on ADHD and attention-related behavioral conditions from relevant new studies.
  - Assess the internal validity (risk of bias) of relevant new studies.
  - Summarize the extent of available new evidence.
  - Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integration, such as study-design heterogeneity.
  - Rate the quality and strength of evidence.

PECO Statement

A PECO (Population, Exposure, Comparator, and Outcome) statement was developed by the review team as an aid to identify search terms and inclusion/exclusion criteria as appropriate for addressing the review question for the systematic review.

Population: Humans without restriction based on age

Exposure:

- PBDE refers to any single PBDE congener or combination of grouped congeners.
- Developmental exposure to PBDEs. To be considered developmental, the exposure occurred during any of the following: prior to conception for one or both parents, during pregnancy (exposure to offspring in utero), perinatally, or in childhood.
- Exposure measurements must be from human biological samples (e.g., urine, blood, or other specimens).

Comparator: Humans exposed to lower levels of PBDEs.

Outcomes:

- Quantitative measures of intelligence. For example, measures from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Wechsler Intelligence Scale for Children (WISC), Stanford-Binet Intelligence Scale, or the McCarthy Scales of Children's Abilities (MSCA).
- Outcome measures of ADHD and attention-related behavioral conditions. For example, measures from the Child Behavior Checklist (CBCL)/1.5-5, Conners' Kiddie Continuous Performance Test (K-CPT), Conners' Rating Scale-Teachers (CRS-T), Conners' Parent Rating Scale-Revised (CPRS), WISC-III (selected subscales), the Disruptive Behavior Disorders Rating Scale (DBD), or Continuous ADHD Confidence Index score.

METHODS

Problem Formulation and Protocol Development

The review question and specific aims were developed and refined through a series of problem formulation steps. The committee considered review articles on endocrine disruptors in surveying the types of chemicals that might make good case examples and held a workshop to explore potential case examples. The committee sought an example of a chemical for which both the human and the animal evidence appears to be associated with different exposure levels of that chemical and due to perturbation of the estrogen or androgen hormone system. PBDEs appear to fit this case criterion. Because the committee learned that other systematic reviews on PBDEs and human neurodevelopment are available, it decided to demonstrate how an existing systematic review can be evaluated for risk of bias and updated.

The protocol will be peer reviewed by subject-matter and systematic-review experts in accordance with standard report-review practices of the National Academies of Sciences, Engineering, and Medicine. The protocols will be revised in response to peer review comments and will subsequently be published as appendices to the committee's final report. The identity of the peer reviewers will remain anonymous to the committee until the publication of the final report, when their names and affiliations are disclosed in the Preface.

Committee and Staff

There are 11 committee members, supported by two staff members of the National Academies. The committee members were appointed in accordance with the standard policies and practices of the National Academies on the basis of their expertise in general toxicology, reproductive toxicology, developmental toxicology, endocrinology, neurotoxicology, epidemiology, risk assessment, biostatistics, and systematic-review methods. The membership of the committee and the staff was determined before the topic of the systematic review was selected. It was known, however, that each case study would be on an endocrine-disrupting chemical, so committee members who have relevant expertise were specifically recruited and appointed.

Review Team

The review team for this case study will be two committee members (KR, AR), two National Academies staff members (EM, SM), and an information specialist (JB). If a member of the review team is found to be a coauthor of a study under review, that member will recuse himself or herself from the evaluation of the quality of that study.

The review team will be responsible for performing all aspects of the review, including conducting the literature searches; applying inclusion/exclusion criteria to screen studies; extracting data; assessing risk of bias for included studies; and analyzing and synthesizing data. The roles and responsibilities of the team members will be documented throughout the protocol. Throughout the course of its work, the review team will also engage other members of the committee to provide consultation as needed. The involvement of those individuals will be documented and acknowledged.

Biographical information on the review team is presented in Section F-1a.

Search Methods

The review team will collaborate with an information specialist (JB) who has training, expertise, and familiarity with developing and performing systematic review literature searches. Recent (within the past 3 years), relevant high-quality systematic reviews addressing the research question about PBDEs and neurodevelopment will be searched. PubMed will be search by adding the qualifier "systematic review"[ti] OR "meta-analysis"[pt] OR "meta-analysis"[ti] OR ("systematic"[ti] AND "review"[ti]) OR (systematic review [tiab] AND review [pt]) OR "meta synthesis"[ti] OR "meta synthesis"[ti] OR "integrative review"[tw] OR "integrative research review"[tw] OR "cochrane database syst rev"[ta] OR "evidence synthesis"[tiab] to the preliminary search strategy (see Section F-1b). Language and date restrictions will be applied (English language; published 2013 to present). The systematic review protocol registry PROSPERO (CRD) will also be searched using key terms from the preliminary PubMed strategy.

Study Selection

Two team members (SM, EM) will independently screen search results, applying the following exclusion criteria:

- Not a systematic review.[1] The minimum criteria for a study to be considered a systematic review are
  - conduct of an explicit and adequate literature search,
  - application of predefined eligibility criteria,

[1]A systematic review "is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (IOM 2011, p. 1).

  - consideration of the quality of included studies or risk of bias assessment, and
  - synthesis (or attempt at synthesis) of the findings, either qualitatively or quantitatively.
- Not in English.
- Search date prior to 2013.
- Does not match the research question or PECO elements.

For PubMed results, screening will be conducted first using abstracts and then at the full-text level. Results from PROSPERO will be conducted at one level, using the information in the registry. Disagreements regarding eligibility will be resolved through discussion or, where necessary, by a third team member.

Assessment for Quality

Eligible systematic reviews of high quality will be reviewed, considering date of search and match with the PECO statement as well as availability of data from the primary studies, how risk of bias was conducted, and other factors. Two investigators (KR, AR) will independently assess the risk of bias of eligible systematic reviews using ROBIS (Whiting et al. 2016). Disagreements in rating will be resolved through discussion or, where necessary, through consultation with a third team member. Systematic reviews rated as low quality will be excluded from further consideration at this stage. Systematic reviews considered a good match will be used to address the research question. Reviews that are a good match but with search dates more than a year ago will be updated.

Updating a Systematic Review

The review team will use the same methods as the existing systematic review to update it.

Search Methods and Study Selection

The review team will update the literature search of the existing review using the strategies from that review and searching from 1 year before the last search date of the review (i.e., an overlap of 1 year). Two team members (SM, EM) will independently apply the same eligibility criteria used in the existing review, first at the title and abstract level and then at the full-text level. A third team member will resolve disagreements, as needed.

Data Extraction and Analysis

The review team will extract data from any newly identified studies into evidence tables with the same structure as in the existing review. Risk of bias will be assessed by two independent team members using the same tool(s) applied in the existing systematic review.

Evidence Synthesis

The review team will qualitatively synthesize the body of evidence for each outcome and, where appropriate, a meta-analysis will be performed. If a meta-analysis is performed, summaries of the main characteristics for each included study will be compiled and reviewed by two team members to determine comparability between studies, to identify data transformations necessary to ensure comparability, and to determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups (e.g., quartiles of exposure)

- Biological measurement for each exposure group
- Health outcome(s) reported
- Conditioning variables in the analysis (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

The review team expects to require input from subject-matter experts to help assess the heterogeneity of the studies. Subgroup analyses to examine the extent to which risk of bias contributes to heterogeneity will be performed. Situations where it may not be appropriate to include a study are when data on exposure or outcome are too different to be combined or other circumstances that may indicate that averaging study results would not produce meaningful results. When considering outcome measures for conducting meta-analyses, continuous outcome measures, such as beta-coefficients (and their associated confidence intervals) from regression analysis, are preferred. A secondary alternative, when there are more than two groups, is to conduct regression analysis of the odds or risk ratios across exposure groups and to use the derived beta coefficient. A tertiary alternative when there are only two groups (e.g., higher and lower exposure) is to use the odds or risk ratio itself.

If a meta-analysis is conducted, a random effects model will be used for the analysis. Heterogeneity will be assessed using the I-squared statistic. Interpretation of I-squared will be based on the Cochrane Handbook: 0% to 40% (might not be important); 30% to 60% (may represent moderate heterogeneity); 50% to 90% (may represent substantial heterogeneity); and 75% to 100% (considerable heterogeneity). Additionally, as described in the Cochrane Handbook, for the last three categories, the importance of the I-squared will be interpreted considering not only the magnitude of effects but also the strength of the evidence (90% two-tailed confidence interval).

The review team will also perform sensitivity analyses on the following aspects:

- Sensitivity to exclusion of individual studies in succession,
- Sensitivity to alternative exposure metrics (if available), and
- Sensitivity to alternative outcome metrics (if available).

It is unlikely that there will be enough studies or information to meaningfully assess publication bias or to perform subgroup analyses, so no such analyses are planned.

In the event that these proposed methods for data analysis are altered to tailor to the evidence base from included studies, the protocol will be amended accordingly and the reasons for change will be justified in the documentation.

Grading/Strength of Evidence

The same system and approach that was used to draw conclusions and grade the evidence in the existing systematic review will be used to characterize the evidence.

REFERENCES

Higgins, J., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration [online]. Available: http://handbook.cocharne.org [accessed May 6, 2016].

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Whiting, P., J. Savovic, J.B. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, and R. Churchill. 2016. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J. Clin. Epidemiol. 69:225-234.

SECTION F-1a

BIOGRAPHICAL INFORMATION ON THE REVIEW TEAM

Jaime F. Blanck is a clinical informationist at the Welch Medical Library at Johns Hopkins University. She creates and implements systematic review search strategies across multiple databases and provides comprehensive reference, research, and information services to multiple departments within the School of Medicine. She received an MLIS from the University of Pittsburgh and an MPA from the University of Baltimore.

Ellen Mantus is a scholar and director of risk assessment on the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine with more than 20 years of experience in the fields of toxicology and risk assessment. She has served as the study director on numerous projects, including ones that have assessed the health implications of various chemical exposures; developed strategies for applying modern scientific approaches in toxicology and risk assessment; provided guidance to federal agencies on risk-based decision making; and evaluated barriers to deployment of electric vehicles and associated charging infrastructure. Before joining the National Academies, Dr. Mantus was a project manager with ICF Consulting where she served as a primary reviewer for numerous toxicological studies and provided risk assessment and regulatory support on a wide array of projects. Dr. Mantus received a PhD in chemistry from Cornell University.

Susan Martel is a senior program officer in the Board on Environmental Studies and Toxicology at the National Academies of Sciences, Engineering, and Medicine. She has 20 years of experience in supporting toxicology and risk assessment projects for the US Environmental Protection Agency, the US Department of Defense, and the National Aeronautics and Space Administration. Recent projects include working with committees evaluating the toxicological effect of arsenic, developing exposure guidelines for use on spacecraft, and assessing pesticide risks-assessment practices. Before joining the National Academies, she was the administrator of the Registry for Toxicology Pathology for Animals at the American Registry of Pathology. She received a BA in biology from Skidmore College.

Karen A. Robinson is an associate professor at the Johns Hopkins University School of Medicine. She also serves as director of the Johns Hopkins University Evidence-based Practice Center and is a member of the core faculty in the Center for Clinical Trials and Evidence Synthesis at the university's Bloomberg School of Public Health. Her research focuses on evidence-based health care and evidence-based research. She conducts systematic reviews that are used to develop clinical practice guidelines and to inform other health decisions. Dr. Robinson received an MSc in health sciences from the University of Waterloo, Ontario, and a PhD in epidemiology from the Johns Hopkins Bloomberg School of Public Health.

Andrew A. Rooney is deputy director of the Office of Health Assessment and Translation (OHAT) in the National Toxicology Program at the National Institute of Environmental Health Sciences. He has been developing risk assessment methods and guidance throughout his professional career and is a principal author of the 2012 WHO/IPCS Guidance for Immunotoxicity Risk Assessment for Chemicals. Most recently, he has been working on emerging issues in toxicology and environmental health, including methods to address study quality in terms of risk of bias for human, animal, and mechanistic studies and adaptation of systematic review methods for addressing environmental health questions. He led the team that developed the OHAT approach to systematic review. Dr. Rooney has an MS and a PhD in zoology from the University of Florida.

SECTION F-1b

The review team will employ a multi-method process to identify all potentially relevant studies as detailed below.

Electronic Searches

PubMed

A search string employing medical subject heading (MeSH) terms and keyword synonyms will be developed. To assist in compiling these terms, the review team will consult an existing systematic review protocol studying PBDEs in humans (J. Lam et al. Applying the navigation guide systematic review methodology. Case study #5: association between developmental exposures to PBDEs and human neurodevelopment. PROSPERO 2015:CRD42015019753 Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42015019753). This protocol was selected because it examines the substances of interests, timing of exposure, and outcomes of interest. The search strategies will address each of the following concepts:

- *Flame retardants (PBDEs)*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to the *Flame retardants (PBDEs)* concept. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. CAS registry numbers for each PBDE substance will also be included in the list of search terms. All MeSH terms, Supplementary Concept terms, keyword synonyms, and CAS registry numbers will be searched together as one concept using the Boolean operator "OR."
- *Human studies*—The search filter developed by the Cochrane Library to identify human studies (see http://handbook.cochrane.org/ part 2, section 6.4.f) will be modified to comply with PubMed formatting.
- *Outcomes*—The review team will use the MeSH database (http://www.ncbi.nlm.nih.gov/mesh) to find all MeSH heading and Supplementary Concept headings that relate to measures of learning, memory, attention, and cognition. The review team will mine the "Entry Terms" list for each of the controlled vocabulary terms identified and include all unique keyword synonyms listed for each. All MeSH terms and keyword synonyms will be searched together as one concept using the Boolean operator "OR."

Each of the above concepts will be searched together using the Boolean operator "AND." There will not be limitations on date of publication, language, or publication type. All citation records will be exported to EndNote. Additional citations identified through the search processes identified below will also be exported to the project EndNote library. Duplicates will be removed from the citation library using the "Find Duplicates" tool in EndNote as well as a manual review of citations by the project librarian to identify any duplicates not found during the automated process. The number of citations found in each database will be recorded, as well as the number of duplicates and final tally of unique citations. The final library of citations will be uploaded to the Health Assessment Workspace Collaboration Web-based tool (www.hawcproject.org) for systematic reviews where they will be reviewed by the team.

Search Strategies

PubMed

("Flame Retardants"[Mesh] OR "Flame Retardants" [Pharmacological Action] OR "Halogenated Diphenyl Ethers"[Mesh] OR "Phenyl Ethers"[Mesh:NoExp] OR "pentabromodiphenyl ether" [Supplementary Con-

cept] OR "2,2',3,3',4,4',6,6'-octabromodiphenyl ether"[Supplementary Concept] OR "decabromobiphenyl ether"[Supplementary Concept] OR "tribromodiphenyl ether 28"[Supplementary Concept] OR "2,2',4,4'-tetrabromodiphenyl ether"[Supplementary Concept] OR "2,2',4,5'-tetrabromodiphenyl ether"[Supplementary Concept] OR "hexabromodiphenyl ether 154"[Supplementary Concept] OR "2,2',4,4',5,6'hexabromodiphenyl ether"[Supplementary Concept] OR "2,2',3,4,4',5',6heptabromodiphenyl ether"[Supplementary Concept] OR "2,2',3,3',4,5,5',6,6'-nonabromodiphenyl ether"[Supplementary Concept] OR "2,2',3,3',4,4',5,6,6'-nonabromodiphenyl ether"[Supplementary Concept] OR "2,2',3,3',4,4',5,5',6-nonabromodiphenyl ether"[Supplementary Concept] OR "2,2',4,4',5,5'-hexabrominated diphenyl ether"[Supplementary Concept] OR "hexabrominated diphenyl ether 153"[Supplementary Concept] OR "pentabrominated diphenyl ether 100"[Supplementary Concept] OR "5-OH-BDE-47"[Supplementary Concept] OR "6-OH-BDE-47"[Supplementary Concept] OR flame retard*[tw] OR fire retard*[tw] OR fireproofing agent*[tw] OR "FireMaster"[tw] OR "Bromkal"[tw] OR diphenyl ether deriv*[tw] OR halogenated diphenyl*[tw] OR brominated diphenyl*[tw] OR PBDE*[tw] OR polybrominated diphenyl*[tw] OR polybromodiphenyl*[tw] OR PBDP*[tw] OR BDE*[tw] OR pentabromodiphenyl*[tw] OR cpentaBDE*[tw] OR PentaBDE*[tw] OR "PeBDE"[tw] OR "DE 71"[tw] OR "DE71"[tw] OR "pentabrominated diphenyl"[tw] OR "pentabrominated diphenyls"[tw] OR "PBDPO"[tw] OR "Planelon PB 501"[tw] OR pentabromo deriv*[tw] OR pentabromophenyl*[tw] OR octabromodiphenyl*[tw] OR c-octaBDE*[tw] OR OctaBDE*[tw] OR "OcBDE"[tw] OR "Octabrom"[tw] OR octabromo deriv*[tw] OR "OBDE"[tw] OR "OBDPO"[tw] OR "octabrominated diphenyl"[tw] OR "octabrominated diphenyls"[tw] OR decabromodiphenyl*[tw] OR cdecaBDE*[tw] OR DecaBDE*[tw] OR "DeBDE"[tw] OR "DBDPO"[tw] OR "decabrominated diphenyl"[tw] OR "decabrominated diphenyls"[tw] OR decabromo deriv*[tw] OR "Decabrom"[tw] OR "Berkflam B 10E"[tw] OR "FR 300BA"[tw] OR "FR 300 BA"[tw] OR tribromodiphenyl*[tw] OR "tribrominated diphenyl"[tw] OR "tribrominated diphenyls"[tw] OR "TrBDE"[tw] OR tribromo deriv*[tw] OR tetrabromodiphenyl*[tw] OR TetraBDE*[tw] OR "TeBDE"[tw] OR "TBDE"[tw] OR "BPDE"[tw] OR tetrabromo deriv*[tw] OR "TBDP"[tw] OR "tetrabrominated diphenyl"[tw] OR "tetrabrominated diphenyls"[tw] OR hexabromodiphenyl*[tw] OR HexaBDE*[tw] OR "HxBDE"[tw] OR "hexabrominated diphenyl"[tw] OR "hexabrominated diphenyls"[tw] OR hexabromo deriv*[tw] OR heptabromodiphenyl*[tw] OR HeptaBDE*[tw] OR "HeBDE"[tw] OR "heptabrominated diphenyl"[tw] OR "heptabrominated diphenyls"[tw] OR heptabromo deriv*[tw] OR nonabromodiphenyl*[tw] OR NonaBDE*[tw] OR "NoBDE"[tw] OR "nonabrominated diphenyl"[tw] OR "nonabrominated diphenyls"[tw] OR nonabromo deriv*[tw] OR "7025-06-1"[tw] OR "6876-00-2"[tw] OR "101-55-3"[tw] OR "51452-870"[tw] OR "446254-14-4"[tw] OR "147217-72-9"[tw] OR "171977-449"[tw] OR "147217-71-8"[tw] OR "33513-66-3"[tw] OR "51930-04-2"[tw] OR "6903-63-5"[tw] OR "189084-59-1"[tw] OR "83694-71-7"[tw] OR "46438-88-4"[tw] OR "2050-47-7"[tw] OR "147217-74-1"[tw] OR "147217-75-2"[tw] OR "407606-55-7"[tw] OR "147217-73-0"[tw] OR "147217-76-3"[tw] OR "337513-67-4"[tw] OR "446254-15-5"[tw] OR "446254-16-6"[tw] OR "147217-77-4"[tw] OR "337513-75-4"[tw] OR "337513-53-8"[tw] OR "41318-75-6"[tw] OR "337513-56-1"[tw] OR "155999-95-4"[tw] OR "65075-08-3"[tw] OR "189084-60-4"[tw] OR "147217-78-5"[tw] OR "446254-17-7"[tw] OR "147217-80-9"[tw] OR "147217-79-6"[tw] OR "147217-81-0"[tw] OR "337513-54-9"[tw] OR "337513-68-5"[tw] OR "446254-18-8"[tw] OR "446254-19-9"[tw] OR "446254-20-2"[tw] OR "446254-22-4"[tw] OR "5436-43-1"[tw] OR "337513-55-0"[tw] OR "243982-82-3"[tw] OR "446254-23-5"[tw] OR "189084-57-9"[tw] OR "446254-24-6"[tw] OR "446254-25-7"[tw] OR "446254-31-5"[tw] OR "446254-32-6"[tw] OR "446254-33-7"[tw] OR "446254-34-8"[tw] OR "189084-61-5"[tw] OR "446254-37-1"[tw] OR "446254-38-2"[tw] OR "327185-09-1"[tw] OR "446254-39-3"[tw] OR "189084-62-6"[tw] OR "446254-40-6"[tw] OR "446254-41-7"[tw] OR "446254-42-8"[tw] OR "189084-63-7"[tw] OR "446254-43-9"[tw] OR "93703-48-1"[tw] OR "446254-45-1"[tw] OR "446254-48-4"[tw] OR "103173-66-6"[tw] OR "446254-50-8"[tw] OR "446254-51-9"[tw] OR "182346-21-0"[tw] OR "446254-53-1"[tw] OR "446254-54-2"[tw] OR "446254-55-3"[tw] OR "446254-55-3"[tw] OR "446254-57-5"[tw] OR "446254-59-7"[tw] OR "446254-61-1"[tw] OR "446254-64-4"[tw] OR "38463-82-0"[tw] OR "60348-60-9"[tw] OR "189084-64-8"[tw] OR "446254-65-5"[tw] OR "446254-66-6"[tw] OR "446254-67-7"[tw] OR "446254-68-8"[tw] OR "373594-78-6"[tw] OR "446254-

69-9"[tw] OR "446254-71-3"[tw] OR "446254-72-4"[tw] OR "446254-74-6"[tw] OR "446254-77-9"[tw] OR "446254-78-0"[tw] OR "189084-65-9"[tw] OR "446254-80-4"[tw] OR "189084-66-0"[tw] OR "182677-30-1"[tw] OR "243982-83-4"[tw] OR "68631-49-2"[tw] OR "207122-15-4"[tw] OR "35854-94-5"[tw] OR "189084-58-0"[tw] OR "189084-67-1"[tw] OR "207122-16-5"[tw] OR "189084-68-2"[tw] OR "1163-19-5"[tw] OR "109945-70-2"[tw] OR "113152-37-7"[tw] OR "113172-79-5"[tw] OR "139598-16-6"[tw] OR "139749-52-3"[tw] OR "145538-74-5"[tw] OR "32534-81-9"[tw] OR "32536-52-0"[tw] OR "40088-47-9"[tw] OR "446254-27-9"[tw] OR "446255-20-5"[tw] OR "446255-22-7"[tw] OR "49690-94-0"[tw] OR "63936-56-1"[tw] OR "64589-00-0"[tw] OR "68928-80-3"[tw] OR "85446-17-9"[tw] OR "36483-60-0"[tw] OR "437701-79-6"[tw] OR "446255-26-1"[tw] OR "117948-63-7"[tw] OR "446255-30-7"[tw] OR "61262-53-1"[tw] OR "405237-85-6"[tw] OR "39275-89-3"[tw] OR "13654-09-6"[tw] OR "61288-13-9"[tw] OR "446255-39-6"[tw] OR "337513-72-1"[tw] OR "366791-32-4"[tw] OR "2050-47-7"[tw]) AND ("Psychological Tests"[Mesh] OR "Mental Disorders Diagnosed in Childhood"[Mesh] OR "Mental Processes"[Mesh] OR "Attention"[Mesh] OR "Human Development"[Mesh] OR "Intelligence"[Mesh] OR "Neurobehavioral Manifestations"[Mesh] OR "Psychomotor Performance"[Mesh] OR "Behavior"[Mesh:NoExp] OR "Adolescent Behavior"[Mesh] OR "Behavioral Symptoms"[Mesh] OR "Child Behavior"[Mesh] OR "Communication"[Mesh] OR "Impulsive Behavior"[Mesh] OR "Motor Activity"[Mesh] OR "Social Behavior"[Mesh] OR "Hyperkinesis"[Mesh] OR "Brain/drug effects"[Mesh] OR "Spatial Behavior"[Mesh] OR neurodevelopment*[tw] OR neurotoxic*[tw] OR neurobehav*[tw] OR neuropsychologic*[tw] OR neurocogniti*[tw] OR psychologic*[tw] OR aptitude*[tw] OR mental*[tw] OR intelligence*[tw] OR "IQ"[tw] OR intellectual*[tw] OR language*[tw] OR comprehension*[tw] OR impulsiv*[tw] OR "ADHD"[tw] OR "ADDH"[tw] OR "ADHS"[tw] OR "AD/HD"[tw] OR "hkd"[tw] OR hyperactiv*[tw] OR hyper activ*[tw] OR hyperkin*[tw] OR hyper kin*[tw] OR attention defic*[tw] OR attention related*[tw] OR inattention*[tw] OR inattentiv*[tw] OR "sustained attention"[tw] OR "attention span"[tw] OR attention dysfunc*[tw] OR attention disorder*[tw] OR "distractibility"[tw] OR Behavioral*[tw] OR behavioural*[tw] OR behavior defic*[tw] OR behaviour defic*[tw] OR behavior dysfunc*[tw] OR behavior disorder*[tw] OR behavior disorder*[tw] OR behavior effect*[tw] OR behaviour effect*[tw] OR behavior checklist*[tw] OR behaviour checklist*[tw] OR disruptive behav*[tw] OR disruption behav*[tw] OR disruptive disorder*[tw] OR disruption disorder*[tw] OR defiance behav*[tw] OR defiant behav*[tw] OR defiance disorder*[tw] OR defiant disorder*[tw] OR spontaneous behav*[tw] OR externalizing behav*[tw] OR "cognitive"[tw] OR "cognition"[tw] OR "psychomotor"[tw] OR "learning"[tw] OR "memory"[tw] OR executive function*[tw] OR executive control*[tw] OR executive dysfunction*[tw] OR executive impairment*[tw] OR motor abilit*[tw] OR motor activit* [tw] OR "motor performance"[tw] OR motor function*[tw] OR motor skill*[tw] OR "fine motor"[tw] OR "vigilance"[tw] OR "reaction time"[tw] OR "processing speed"[tw] OR "response inhibition"[tw] OR "Stanford Binet"[tw] OR Binet Test*[tw] OR "Bender Gestalt Test"[tw] OR Aphasia Test*[tw] OR Bayley*[tw] OR "Wechsler"[tw] OR "WISC"[tw] OR McCarthy Scale*[tw] OR "Continuous Performance Test"[tw] OR "Continuous Performance Tests"[tw] OR "Continuous Performance Task"[tw] OR "Continuous Performance Tasks"[tw] OR Conners*[tw] OR "CRS-T"[tw] OR "CRS-P"[tw] OR "academic achievement"[tw] OR "scholastic achievement"[tw] OR brain disorder*[tw] OR brain damage*[tw] OR brain dysfunct*[tw]) AND (("systematic review"[ti] OR "meta-analysis"[pt] OR "meta-analysis"[ti] OR ("systematic"[ti] AND "review"[ti]) OR (systematic review [tiab] AND review [pt]) OR "meta synthesis"[ti] OR "meta synthesis"[ti] OR "integrative review"[tw] OR "integrative research review"[tw] OR "cochrane database syst rev"[ta] OR "evidence synthesis"[tiab])) NOT ((("Animals"[Mesh] NOT ("Animals"[Mesh] AND ("Humans"[Mesh]))))

SECTION F-1c

AMENDMENTS TO THE PROTOCOL
(November 11, 2016)

Additions to the Review Team

The following committee members were added to the review team to supplement expertise:

- **David C. Dorman** (*Chair*) is a professor of toxicology in the Department of Molecular Biosciences of North Carolina State University. The primary objective of his research is to provide a refined understanding of chemically induced neurotoxicity in laboratory animals that will lead to improved assessment of potential toxicity in humans. Dr. Dorman's research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has chaired or served on several NRC committees, including the Committee on Design and Evaluation of Safer Chemical Substitutions: A Framework to Inform Government and Industry Decisions, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He has served on other advisory boards for the US Navy, NASA, and USDA and is currently a member of NTP's Board of Scientific Counselors. Dr. Dorman is an elected fellow of the Academy of Toxicological Sciences and a fellow of the American Association for the Advancement of Sciences. He received a DVM from Colorado State University. He completed a combined PhD and residency program in toxicology at the University of Illinois at Urbana-Champaign, and is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.
- **Barbara F. Hales** is a James McGill Professor in the Department of Pharmacology and Therapeutics at McGill University. Her research interests are in the mechanisms of action of drugs as teratogens. She studies developmental toxicity using a combination of in vivo, in vitro, and molecular approaches with the goal of elucidating how the embryo responds to insult after direct or maternal exposure and the consequences to progeny of paternal drug exposure. Dr. Hales is a past president of the Teratology Society and is currently co-chair of the Chemicals Management Plan Science Committee of the Government of Canada. She received an MSc in pharmacognosy from the Philadelphia College of Pharmacy and Science and a PhD in pharmacology and therapeutics from McGill University.
- **Susan L. Schantz** is a professor of toxicology in the Department of Comparative Biosciences, College of Veterinary Medicine, at the University of Illinois at Urbana-Champaign. She is also director of a National Institute of Environmental Health Sciences (NIEHS) T32 training program in endocrine, developmental, and reproductive toxicology and director of a Children's Environmental Health Research Center jointly funded by the NIEHS and the EPA. In addition, she is currently the interim director of the Neuroscience Program. Dr. Schantz's research interests involve understanding the neurobehavioral effects of chemical exposures during development and aging. She conducts research in both laboratory-based animal studies and parallel epidemiologic studies. She has served as president of the Neurotoxicology Specialty Section of the Society of Toxicology and president of the Neurobehavioral Teratology Society. Dr. Schantz was also a member of the NRC's Committee to Assess the Health Implications of Perchlorate Ingestion. She received a PhD in environmental toxicology from the University of Wisconsin–Madison.

SECTION F-2

Results of Literature Searches for Existing or Ongoing Systematic Reviews

Searches for systematic reviews or ongoing reviews were performed on August 3, 2016. Five publications were found in PubMed and 13 protocols in PROSPERO. Below is the list of the18 reports.

Ahmed, I., E. Dickenson, A. Sprowson, and N. Parsons. 2014. The Use of Triclosan Coated Sutures to Prevent Surgical Site Infections: A Systematic Review and Meta-Analysis of the Literature. PROSPERO 2014: CRD42014014856 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014856 [accessed August 3, 2016].

Bonde, J.P., E. Bräuner, I.O. Sprecht, C. Glazer, K.K. Hærvig, S.E. Bondo Petersen, E. Flaches, B. Høyer, L. Rylander, S. Andersen, K.S. Hougaard, G. Toft, C. Ramlau-Hansen, L. Rylander, A. Giwercman, and S. Andersen. 2016. The Epidemiologic Evidence Linking Pre- and Postnatal Exposure to Endocrine Disrupting Chemicals with Male Reproductive Disorders: A Systematic Review and Meta-analysis. PROSPERO 2016:CRD42016037427 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016037427 [accessed August 3, 2016].

Bramwell, L., S.V. Glinianaia, J. Rankin, M. Rose, A. Fernandes, S. Harrad, and T. Pless-Mulolli. 2016. Associations between human exposure to polybrominated diphenyl ether flame retardants via diet and indoor dust, and internal dose: A systematic review. Environ. Int. 93:680-694.

de Sousa, A.T., N.S. Formiga, S.H. Oliveira, M.M. Costa, and M.J. Soares. 2015. Using the theory of meaningful learning in nursing education. Rev. Bras. Enferm. 68(4):626-635.

Grandjean, P., and P.J. Landrigan. 2014. Neurobehavioral effects of developmental toxicity. Lancet Neurol. 13(3):330-338.

Kim, Y.R., F.A. Harden, L.M. Toms, and R.E. Norman. 2014. Health consequences of exposure to brominated flame retardants: A systematic review. Chemosphere 106:1-19.

Lam, J., P. Sutton, J. McPartland, L.I. Davidson, N. Daniels, S. Sen, D. Axelrad, B. Lanphear, D. Bellinger, and T.J. Woodruff. 2015. Applying the Navigation Guide Systematic Review Methodology, Case Study No. 5. Association between Developmental Exposures to PBDEs and Human Neurodevelopment: A Systematic Review of the Evidence Protocol. PROSPERO 2015:CRD42015019753 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015019753 [accessed August 3, 2016].

Líbera, B.D., P.A. Ribeiro Neves, C. Saunders, M. Baião, and D. Cavalcante Barros. 2013. The Role of Prenatal Nutritional Assistance in the Context of Primary Care: A Systematic Review. PROSPERO 2013:CRD42013005389 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013005389 [accessed August 3, 2016].

Lotta, L.A., A. Abbasi, A.S. Shalqvist, J. Wilk, D. Nunez, J. Brosnan, D. Waterworth, and N. Wareham. 2014. Effect of Fibrates on Metabolic Traits in Non-diabetic Individuals: A Meta-analysis of Randomized Controlled Trials. PROSPERO 2014:CRD42014013683 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014013683 [accessed August 3, 2016].

Marcolino, M., L. Maia, B. Pereira, J. Oliveira, D. Andrade-Junior, A. Ribeiro, and E. Boersma. 2016. Impact of Telemedicine Interventions on Time to Reperfusion and Mortality in Acute Myocardial Infarction Patients. PROSPERO 2016:CRD42016025404 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016025404 [accessed August 3, 2016].

Mozetic, V. 2016. Statins and Fibrates for Diabetic Retinopathy: Systematic Review. PROSPERO 2016:CRD42016029746 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016029746 [accessed August 3, 2016].

Pase, C., C. Nazareth, R. Almeida, and C. Garcia Rodrigues. 2016. How the Healthcare Team Deals with The Loss of Pediatric Patient: A Systematic Review with Meta-Summarization. PROSPERO 2016:CRD42016032698 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013005389 [accessed August 3, 2016].

Roth, N., and M.F. Wilks. 2014. Neurodevelopmental and neurobehavioral effects of polybrominated and perfluorinated chemicals: A systematic review of the epidemiological literature using a quality assessment scheme. Toxicol. Lett. 230(2):271-281.

Sadeghirad, B., J. Erickson, L. Lytvyn, T. Webber-Adams, J. Slavin, and B. Johnston. 2015. Scientific Basis for Recommendations on Sugars From Authoritative Health Organizations: A Systematic Review of Public Health Guidelines. PROSPERO 2015:CRD42015029182 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015029182 [accessed August 3, 2016].

Sugeng, E., M. de Cock, and M. van de Bor. 2016. Toddler Exposure to Flame Retardant Chemicals: Magnitude, Health Concern and Potential Sources of Exposure: Observational Studies Summarized in a Systematic Review. PROSPERO 2016:CRD42016043245 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016043245 [accessed August 3, 2016].

Tarp, S., E.M. Bartels, H. Bliddal, B. Danneskiold,-Samsøe, M. Rasmussen, and R. Christensen. 2011. Effect of Non-steroidal Anti-inflammatory Drugs on C-reactive Protein Levels in Patients with Rheumatoid Arthritis: A Meta-analysis of Randomized, Double Blind, Placebo-controlled Trials. PROSPERO 2011:CRD 42011001157 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001157 [accessed August 3, 2016].

Zhao, X.M. 2016. Hair as Noninvasive Biomarkers of Human Exposure to Polybrominated Diphenyl Ethers (PBDEs). PROSPERO 2016:CRD42016039582 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016039582 [accessed August 3, 2016].

Zhao, X.M., J. Li, and H.L. Wang. 2014. The Correlations Between Polybrominated Diphenyl Ethers (PBDEs) and Thyroid Hormones in the General Population: A Meta-analysis. PROSPERO 2014:CRD42014013289 [online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014013289 [accessed August 3, 2016].

SECTION F-3

Results of Literature Searches for Human Studies on the Effects of Developmental Exposure to PBDEs on Intelligence or ADHD and Attention-Related Behavioral Conditions

Literature searches were performed on September 28, 2016, using the search strategy presented in Lam et al. (2015). The search was restricted to reports published after March 5, 2014, so that the search would have a 1-year overlap with the ongoing Lam et al. review. A summary of the results is presented below.

BIOSIS:	75
Embase:	291
PubMed:	179
ToxNet/DART:	265
Web of Science:	141
Other sources:	29

Total citations found:	980
Duplicates removed:	523
Total unique citations:	457

SECTION F-4

Human Studies of PBDES and Intelligence and ADHD and Attention-Related Behavioral Conditions

Studies in the Lam et al. (2015) Systematic Review and Three New Reports (highlighted in table)

Reference	Population	Location	Sample Size	PBDE Measured	PBDE Concentration	Exposure Matrix			Outcome Assessment
						Cord/Maternal Blood	Child Serum/ Blood	Breast Milk	
IQ Measures Only									
Prospective Birth Cohort									
Gascon et al. 2012	Pregnant women enrolled in INfancia y Medio Ambiente (INMA) (Environment and Childhood) Project between 2004 and 2008	Gipuzkoa, Basque Country, and Sabadell, Catalonia, Spain	290 mother-child pairs	47, 99, 100, 153,154,183, 209, and sum of all seven	BDE 47 range: <LOD-5 ng/g lipid			X	Bayley (BSID) mental score at 12-18 months.
Lin et al. 2010	Pregnant women randomly recruited from four local hospitals between 2007 and 2008	Southern Taiwan	35 mother-child pairs	47, 99, 100, 153, 154, 196, 197, 206, 207, 208, 209, and sum of all 11	Mean of BDE sum: 7.00 ng/g lipid; median = 2.50			X	Bayley-III Cognitive, Bayley-III Language assessed at 8-13 months.
Herbstman et al. 2010	Women pregnant on September 11, 2001, who subsequently delivered babies in one of three downtown hospitals (Beth Israel, St. Vincent's, and St. Vincent's affiliated Elizabeth Seton Childbearing Center)	New York City, NY, USA	152 mother-child pairs	47, 99, 100, 153	BDE 47 range: <LOD-613.1 ng/g lipid	X			Bayley-II Mental Development Index (MDI), Bayley-II Psychomotor Development Index (PDI), and Wechsler Preschool and Primary Scale of Intelligence, Revised Edition (WPPSI-R) Full Scale IQ. Bayley measured at 12, 24, and 36 months. WPPSI-R measured at 48 and 72 months.
Shy et al. 2011	Pregnant women randomly recruited from four local hospitals between 2007 and 2008	Southern Taiwan	36 mother-child pairs	15, 28, 47, 49, 99, 100, 153, 154, 183, 196, 197, and sum of all 11	BDE-47 range: 0.351-19.6 ng/g lipid	X			Bayley III Cognitive, Language subscale assessed at 8-12 months.
Post-hoc Analysis of Prospective Birth Cohort									
Chao et al. 2011	Pregnant women randomly recruited from four local hospitals between 2007 and 2010	Southern Taiwan	70 mother-child pairs	28, 47, 99, 100, 153, 154, 183, 196, 197, 203, 206, 207, 208, 209, and sum of all 14	BDE 47 range: 0.207-80.4 ng/g lipid			X	Bayley-III Cognitive, Bayley-III Language assessed at 8-12 months.

Prospective Birth Cohort									
Adgent et al. 2014	Pregnant women enrolled in the Pregnancy, Infection, and Nutrition (PIN) Babies Study between 2004 and 2006	Central North Carolina, USA	304 mother-child pairs	28, 47, 99, 100, and 153	Median: 27.7 ng/g lipid (IQR: 15.7, 54.2)			X	Mullen Scales of Early Learning composite score and Behavioral Assessment System for Children 2 (BASC-2) (attention subscale) (n = 192) measured at 36 months.
Chen et al. 2014	Pregnant women enrolled in the Health Outcomes and Measures of the Environment (HOME) Study between 2003 and 2006	Cincinnati, Ohio, USA	309 mother-infant pairs	47 and sum of 47, 99, 100, and 153	BDE sum 10th-90th percentile range: 6.4-67.9 ng/g lipid	X			Mental development index assessed by Bayley Scales of Infant Development-II (BSID-II) at 12, 24, and 36 months; Full scale IQ assessed by Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III) at 60 months. Attention/hyperactivity assessed by Behavioral Assessment System for Children-2 (BASC-2) at 2, 3, 4, and 5 years.
Zhang et al. 2017			239 mother-infant pairs		Median BDE sum: 35.65 ng/g lipid				Wechsler Intelligence Scale for Children-IV (WISC-IV) to obtain full scale IQ and Behavioral Assessment System for Children-2 (BASC-2) at 8 years.
Eskenazi et al. 2013	CHAM1: Pregnant women enrolled in the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) study	Salinas Valley, California, USA	212 to 266 (depending on outcome assessed) mother-child pairs	17, 28, 47, 66, 85, 99, 100, 153, 154, 183, sum of 47, 99, 100, and 153, and sum of all 10	BDE 47 range: <LOD-761 ng/g lipid	X	X		Performance IQ at 60 months, full scale IQ at 7 years; CBCL attention problems, CBCL ADHD, K-CPT ADHD Confidence Index, Conners' Rating Scale maternal report ADHD index and DSM-IV inattentive and hyperactivity/impulsivity subscales, BASC-2 maternal report hyperactivity scale and attention problems scale, Conners' rating scale teacher report ADHD index and DSM-IV inattentive and hyperactivity/impusivity subscales, BASC-2 teacher report hyperactivity scale and attention problems scale at 5 and 7 years.

Sagiv et al. 2015	CHAM2: additional children recruited between 2009 and 2011		CHAM1: 321 children; CHAM2: 301 children		CHAM2: Prenatal exposure values estimated by back extrapolation Geometric mean of BDE 47 (CHAM 1 and 2): 15.6 ng/g lipid			Conners' Continuous Performance Test (CPT II), Wechsler Intelligence Scale for Children (WISC-IV), Conners' ADHD-DSM-IV Scales, Parent Versions (CADS-P), Behavior Assessment System for Children, 2nd edition, Parent Report (BASC-2) and Self-Report of Personality (SRP).
Gascon et al. 2011	Pregnant women enrolled in INfancia y Medio Ambiente (INMA) (Environment and Childhood) Project between 1997 and 2001	Gipuzkoa, Basque Country, and Sabadell, Catalonia, Spain	Cord blood: 47 mother-infant pairs; serum: 240 mother-infant pairs	47	BDE 47 range: <LOQ-16.8 ng/g lipid	X		McCarthy Scales of Children's Abilities (MSCA) total cognitive function score measured at 48 months. ADHD-DSM-IV for attention deficit and hyperactivity measured at 4 years.
Attention Measures Only								
Prospective Birth Cohort								
Hoffman et al. 2012	Pregnant women enrolled in the Pregnancy, Infection, and Nutrition (PIN) Babies Study between 2001 and 2005	Central North Carolina, USA	222 mother-child pairs	28, 47, 99, 100, 153, and sum of all five	BDE 47 range: 4-1,430 ng/g lipid		X	Infant-Toddler Social and Emotional Assessment (ITSEA)-activity/impulsivity and attention subscales measured at 24-36 months.
Roze et al. 2009	Pregnant women in Groningen Infant COMPARE (Comparison of the Exposure-Effect Pathways to Improve the Assessment of Human Health Risk of Complex Environmental Mixtures of Organohalogens) (GIC) 2001-2007	Northern providences of the Netherlands	62 mother-child pairs	47, 99, 100	BDE 47 range: <LOD-6.1 ng/g lipid	X		Child Behavior Checklist (CBCL)—attention sustained and attention selective subscales assessed at 5-6 years.
Cowell et al. 2015	Women pregnant on September 11, 2001, who subsequently delivered babies in one of three downtown hospitals (Beth Israel, St. Vincent's, and St. Vincent's affiliated Elizabeth Seton Childbearing Center)	New York City, NY, USA	109 children at age 4; 107 children at age 6	47, 99, 100, 153	Age 4: median BDE 47 = 12.0 ng/g lipid; Age 6: median BDE 47 = 11.4 ng/g lipid	X		Child Behavior Checklist annually at 3-7 years.

Cross-Sectional Study							
Gump et al. 2014	Children recruited from another ongoing study regarding effects of lead	Oswego County, New York, USA	43 children	28, 47, 99, 100	BDE 47 range: <LOQ-0.378 ng/g lipid	X	Parental Strengths and Difficulties Questionnaire (SDQ) hyperactivity-inattention subscale at 10 years.

REFERENCES

Adgent, M.A., K. Hoffman, B.D. Goldman, A. Sjödin, and J.L. Daniels. 2014. Brominated flame retardants in breast milk and behavioural and cognitive development at 36 months. Paediatr. Perinat. Epidemiol. 28(1):48-57.

Chao, R.H., T.C. Tsou, H.L. Huang, and G.P. Chang-Chien. 2011. Levels of breast milk PBDEs from southern Taiwan and their potential impact on neurodevelopment. Pediatr. Res. 70(6):596-600.

Chen, A., K. Yolton, S.A. Rauch, G.M. Webster, R. Hornung, A. Sjödin, K.N. Dietrich, and B.P. Lanphear. 2014. Prenatal polybrominated diphenyl ether exposures and neurodevelopment in U.S. children through 5 years of age: The HOME study. Environ. Health Perspect. 122(8):856-862.

Cowell, W.J., S.A. Lederman, A. Sjödin, R. Jones, S. Wang, F.P. Perera, R. Wang, V.A. Rauh, and J.B. Herbstman. 2015. Prenatal exposure to polybrominated diphenyl ethers and child attention problems at 3-7 years. Neurotoxicol. Teratol. 52(Pt B):143-150.

Eskenazi, B., J. Chevrier, S.A. Rauch, K. Kogut, K.G. Harley, C. Johnson, C. Trujillo, A. Sjödin, and A. Bradman. 2013. In utero and childhood polybrominated diphenyl ether (PBDE) exposures and neurodevelopment in the CHAMACOS study. Environ. Health Perspect. 121(2):257-262.

Gascon, M., M. Vrijheid, D. Martínez, J. Forns, J.O. Grimalt, M. Torrent, and J. Sunyer. 2011. Effects of pre and postnatal exposure to low levels of polybromodiphenyl ethers on neurodevelopment and thyroid hormone levels at 4 years of age. Environ. Int. 37(3):605-611.

Gascon, M., M. Fort, D. Martínez, A.E. Carsin, J. Forns, J.O. Grimalt, L. Santa Marina, N. Lertxundi, J. Sunyer, and M. Vrijheid. 2012. Polybrominated diphenyl ethers (PBDEs) in breast milk and neuropsychological development in infants. Environ. Health Perspect. 120(12):1760-1765.

Gump, B.B., S. Yun, and K. Kannan. 2014. Polybrominated diphenyl ether (PBDE) exposure in children: Possible associations with cardiovascular and psychological functions. Environ. Res.132:244-250.

Herbstman, J.B., A. Sjödin, M. Kurzon, S.A. Lederman, R.S. Jones, V. Rauh, L.L. Needham, D. Tang, M. Niedzwiecki, R.Y. Wang, and F. Perera. 2010. Prenatal exposure to PBDEs and neurodevelopment. Environ. Health Perspect. 118(5):712-719.

Hoffman, K., M. Adgent, B.D. Goldman, A. Sjödin, and J.L. Daniels. 2012. Lactational exposure to polybrominated diphenyl ethers and its relation to social and emotional development among toddlers. Environ. Health Perspect. 120(10):1438-1442.

Lin, D.Y., H.R. Chao, Y.Y. Gou, and C.Y. Huang. 2010. Infants ingesting high breast milk levels of polybrominated diphenyl ethers may have negative impact on their neurodevelopment. Pp. 325-327 in 2010 International Conference on Chemistry and Chemical Engineering (ICCCE 2010) IEEE [online]. Available: http://ieeexplore.ieee.org/document/5560424/#full-text-section [accessed November 10, 2016].

Roze, E., L. Meijer, A. Bakker, K.N. Van Braeckel, P.J. Sauer, and A.F. Bos. 2009. Prenatal exposure to organohalogens, including brominated flame retardants, influences motor, cognitive, and behavioral performance at school age. Environ. Health Perspect. 117(12):1953-1958.

Sagiv, S.K., K. Kogut, F.W. Gaspar, R.B. Gunier, K.G. Harley, K. Parra, D. Villasenor, A. Bradman, N. Holland, and B. Eskenazi**.** 2015. Prenatal and childhood polybrominated diphenyl ether (PBDE) exposure and attention and executive function at 9-12 years of age. Neurotoxicol. Teratol. 52(Pt B):151-161.

Shy, C.G., H.L. Huang, G.P. Chang-Chien, H.R. Chao, and T.C. Tsou. 2011. Neurodevelopment of infants with prenatal exposure to polybrominated diphenyl ethers. Bull. Environ. Contam. Toxicol. 87(6):643-648.

Zhang, H., K. Yolton, G.M. Webster, A. Sjodin, A.M. Calafat, K.N. Dietrich, Y. Xu, C. Xie, J.M. Braun, B.P. Lanphear, and A. Chen. 2017. Prenatal PBDE and PCB exposures and reading, cognition, and externalizing behavior in children. Environ. Health Perspect.