Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose-Response Relationships as they Apply to Endocrine Disrupters

ISBN
978-0-309-29754-7

66 pages
6 x 9
PAPERBACK (2014)

Committee to Review EPAâ€™s State of the Science Paper on Nonmonotonic Dose Reponse; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council

REVIEW OF THE ENVIRONMENTAL PROTECTION AGENCY'S STATE-OF-THE-SCIENCE EVALUATION OF NONMONOTONIC DOSE-RESPONSE RELATIONSHIPS AS THEY APPLY TO ENDOCRINE DISRUPTORS

Committee to Review EPA's State of the Science Paper
on Nonmonotonic Dose Response

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

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

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This project was supported by Contract EP-C-09-003 between the National Academy of Sciences and the US Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-29754-7
International Standard Book Number-10: 0-309-29754-0

Additional copies of this report 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

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE TO REVIEW EPA'S DRAFT STATE OF THE SCIENCE PAPER ON NONMONOTONIC DOSE RESPONSE

Members

DAVID A. SAVITZ (*Chair*), Brown University, Providence, RI
ANDREA BACCARELLI, Harvard School of Public Health, Boston, MA
ROBERT E. CHAPIN, Pfizer, Inc., Groton, CT
RICHARD A. CORLEY, Pacific Northwest National Laboratory, Richland, WA
GEORGE P. DASTON, Procter and Gamble Company, Mason, OH
RUSS B. HAUSER, Harvard School of Public Health, Boston, MA
AMY H. HERRING, University of North Carolina, Chapel Hill, NC
ANDREAS KORTENKAMP, Brunel University, Uxbridge, United Kingdom
HEATHER B. PATISAUL, North Carolina State University, Raleigh, NC
ELIZABETH N. PEARCE, Boston University, Boston, MA
TRACEY J. WOODRUFF, University of California San Francisco, CA
LAUREN ZEISE, California Environmental Protection Agency, Oakland, CA
YILIANG ZHU, University of South Florida, Tampa, FL

Staff

SUSAN N.J. MARTEL, Project Director
KERI STOEVER, Research Associate
NORMAN GROSSBLATT, Senior Editor
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
RADIAH ROSE, Manager, Editorial Projects
TAMARA DAWSON, Program Associate

Sponsor

US Environmental Protection Agency

OTHER REPORTS OF THE
BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Review of EPA's Integrated Risk Information System (IRIS) Process (2014)
Assessing Risks to Endangered and Threatened Species from Pesticides (2013)
Science for Environmental Protection: The Road Ahead (2012)
Exposure Science in the 21st Century: A Vision and A Strategy (2012)
A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials (2012)
Macondo Well–Deepwater Horizon Blowout: Lessons for Improving Offshore Drilling Safety (2012)
Feasibility of Using Mycoherbicides for Controlling Illicit Drug Crops (2011)
Improving Health in the United States: The Role of Health Impact Assessment (2011)
A Risk-Characterization Framework for Decision-Making at the Food and Drug Administration (2011)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (2011)
Toxicity-Pathway-Based Risk Assessment: Preparing for Paradigm Change (2010)
The Use of Title 42 Authority at the U.S. Environmental Protection Agency (2010)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene (2010)
Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use (2009)
Contaminated Water Supplies at Camp Lejeune—Assessing Potential Health Effects (2009)
Review of the Federal Strategy for Nanotechnology-Related Environmental, Health, and Safety Research (2009)
Science and Decisions: Advancing Risk Assessment (2009)
Phthalates and Cumulative Risk Assessment: The Tasks Ahead (2008)
Estimating Mortality Risk Reduction and Economic Benefits from Controlling Ozone Air Pollution (2008)
Respiratory Diseases Research at NIOSH (2008)
Evaluating Research Efficiency in the U.S. Environmental Protection Agency (2008)
Hydrology, Ecology, and Fishes of the Klamath River Basin (2008)
Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment (2007)
Models in Environmental Regulatory Decision Making (2007)
Toxicity Testing in the Twenty-first Century: A Vision and a Strategy (2007)
Sediment Dredging at Superfund Megasites: Assessing the Effectiveness (2007)
Environmental Impacts of Wind-Energy Projects (2007)
Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget (2007)
Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues (2006)
New Source Review for Stationary Sources of Air Pollution (2006)
Human Biomonitoring for Environmental Chemicals (2006)
Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment (2006)

Fluoride in Drinking Water: A Scientific Review of EPA's Standards (2006)
State and Federal Standards for Mobile-Source Emissions (2006)
Superfund and Mining Megasites—Lessons from the Coeur d'Alene River Basin (2005)
Health Implications of Perchlorate Ingestion (2005)
Air Quality Management in the United States (2004)
Endangered and Threatened Species of the Platte River (2004)
Atlantic Salmon in Maine (2004)
Endangered and Threatened Fishes in the Klamath River Basin (2004)
Cumulative Environmental Effects of Alaska North Slope Oil and Gas Development (2003)
Estimating the Public Health Benefits of Proposed Air Pollution Regulations (2002)
Biosolids Applied to Land: Advancing Standards and Practices (2002)
The Airliner Cabin Environment and Health of Passengers and Crew (2002)
Arsenic in Drinking Water: 2001 Update (2001)
Evaluating Vehicle Emissions Inspection and Maintenance Programs (2001)
Compensating for Wetland Losses Under the Clean Water Act (2001)
A Risk-Management Strategy for PCB-Contaminated Sediments (2001)
Acute Exposure Guideline Levels for Selected Airborne Chemicals (seventeen volumes, 2000-2014)
Toxicological Effects of Methylmercury (2000)
Strengthening Science at the U.S. Environmental Protection Agency (2000)
Scientific Frontiers in Developmental Toxicology and Risk Assessment (2000)
Ecological Indicators for the Nation (2000)
Waste Incineration and Public Health (2000)
Hormonally Active Agents in the Environment (1999)
Research Priorities for Airborne Particulate Matter (four volumes, 1998-2004)
The National Research Council's Committee on Toxicology: The First 50 Years (1997)
Carcinogens and Anticarcinogens in the Human Diet (1996)
Upstream: Salmon and Society in the Pacific Northwest (1996)
Science and the Endangered Species Act (1995)
Wetlands: Characteristics and Boundaries (1995)
Biologic Markers (five volumes, 1989-1995)
Science and Judgment in Risk Assessment (1994)
Pesticides in the Diets of Infants and Children (1993)
Dolphins and the Tuna Industry (1992)
Science and the National Parks (1992)
Human Exposure Assessment for Airborne Pollutants (1991)
Rethinking the Ozone Problem in Urban and Regional Air Pollution (1991)
Decline of the Sea Turtles (1990)

Copies of these reports may be ordered from the National Academies Press
(800) 624-6242 or (202) 334-3313
www.nap.edu

Preface

Potential health effects of endocrine-disrupting chemicals raise an environmental health concern about the chemicals' ability to interfere with normal hormone function in humans and wildlife. Some research on endocrine-disrupting chemicals has identified dose–response relationships that have nonmonotonic curves; that is, lower doses are not always associated with smaller responses, nor higher doses with larger responses. Nonmonotonic dose–response (NMDR) curves have been a subject of debate in regulatory toxicology because of their implications for how chemicals should be tested and for how risks posed by such chemicals should be assessed. The debate has focused on whether standard toxicity-testing protocols assess relevant health effects for such chemicals, on the degree to which such tests might miss low-dose effects of NMDR relationships (false-negative results), and on the implications of the evidence on NMDR curves for current risk-assessment practices. To help to address those issues, the US Environmental Protection Agency (EPA) developed a draft *State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures*. EPA asked the National Research Council to conduct an independent review of its evaluation to ensure that it is scientifically sound and of high quality.

In response to EPA's request, the National Research Council convened the Committee to Review EPA's Draft State of the Science Paper on Nonmonotonic Dose Response, which prepared this report. The members of the committee were selected for their expertise in reproductive and developmental toxicology, endocrinology, epidemiology, environmental epigenetics, toxicogenomics, mechanistic toxicology, physiologically based pharmacokinetic modeling, risk assessment, and biostatistics. Biographic information on the committee members is presented in Appendix A.

The committee's report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of the independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards of 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 for their review of the report: Sandra Baird, Massachusetts Department of Environmental Protection; Niladri Basu, McGill University; Jan-Åke Gustafsson, University of Houston; Sangtae Kim, Purdue University; M. Sue Marty, The Dow Chemical Company; Shyamal Peddada, National Institute of Environmental Health Sciences; Gail Prins, University of Illinois at Chicago; Justin Teeguarden, Pacific Northwest National Laboratory; Rochelle Tyl, RTI International; John Vandenbergh, North Carolina State University; and R. Thomas Zoeller, University of Massachusetts Amherst.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by Linda McCauley, Emory University, and Martin Philbert, University of Michigan. Appointed by the National Research Council, they were responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the report rests entirely with the author committee and the institution.

The committee is grateful for the assistance of National Research Council staff in preparing the report. It particularly wishes to acknowledge with deep gratitude the support of Project Director Susan Martel, who coordinated the project and contributed to the committee's report. Other staff members who contributed to this effort are James Reisa, director of the Board on Environmental Studies and Toxicology; Keri Stoever, research associate; Tamara Dawson, program associate; Norman Grossblatt, senior editor; and Mirsada Karalic-Loncarevic, manager of the Technical Information Center.

Finally, I thank all the members of the committee for their efforts throughout the development of this report. Their varied expertise attests to the complexity of this topic, and the committee's ability to synthesize that expertise into a clear consensus is much appreciated.

David A. Savitz, PhD, *Chair*
Committee to Review EPA's Draft State
of the Science Paper on Nonmonotonic
Dose Response

Contents

FIGURES, TABLES, AND BOXES

FIGURES

TABLE

BOXES

REVIEW OF THE ENVIRONMENTAL PROTECTION AGENCY'S STATE-OF-THE-SCIENCE EVALUATION OF NONMONOTONIC DOSE-RESPONSE RELATIONSHIPS AS THEY APPLY TO ENDOCRINE DISRUPTORS

Summary

Potential health effects of chemicals that disrupt endocrine function pose an environmental health concern about their ability to interfere with normal hormone function in human and wildlife populations. The endocrine system regulates biologic processes throughout the body and can be sensitive to small changes in hormone concentrations. Endocrine-disruptor research has focused primarily on chemicals that affect three hormone pathways that play important roles in reproduction and development—the estrogen, androgen, and thyroid hormone pathways. Some of the research has identified dose–response relationships that have nonmonotonic curves. Nonmonotonic dose–response (NMDR) curves are of concern because they do not follow the usual assumption made in toxicology that toxic response decreases as dose decreases. The slope of a NMDR curve changes sign, and the function can take on the shape of a U, the shape of an inverted U, or another shape that has more than one inflection point.

The existence of NMDR curves has been controversial for decades, and there has been considerable debate about their implications for the testing of chemicals and the assessment of risks posed by chemicals. Toxicity tests are designed to identify hazards and to characterize dose–response relationships, so they are aimed at finding a dose that is high enough to elicit a response and exploring dose–response relationships by spacing lower doses to identify a no-observed-adverse-effect level (NOAEL) or a lowest observed-adverse-effect level. One concern raised by NMDR relationships is that such studies as currently designed might not detect critical points (such as peaks and valleys) along a dose–response curve if only a few doses are tested or if the inflection point occurs below the doses tested. Another concern is that some NMDR relationships are found in connection with biologic effects that are not usually evaluated in toxicity tests. If current testing strategies are inadequate to account for NMDR relationships, changes in risk-assessment practices might be necessary.

To address these concerns, the US Environmental Protection Agency (EPA) has developed 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* (SOTS evaluation). The SOTS evaluation presents a collection of the evidence on NMDR relationships to answer three central scientific questions:

1. Do [NMDRs] exist for chemicals and if so under what conditions do they occur?
2. Do NMDRs capture adverse effects that are not captured using [EPA's] current chemical testing strategies (i.e., false negatives)?
3. Do NMDRs provide key information that would alter EPA's current weight of evidence [WOE] conclusions and risk assessment determinations, either qualitatively or quantitatively?

Given the important role that the SOTS evaluation could play in making decisions about conventional toxicity testing and risk-assessment practices, EPA asked the National Research Council to evaluate whether it presents a scientifically sound and high-quality analysis of the literature on NMDR curves. To the committee's knowledge, this is one of the few attempts to evaluate specifically the implications of the evidence on NMDR curves for testing strategies and risk-assessment practices, and the agency is commended for undertaking such a challenging task.

This report presents the findings of the committee convened by the National Research Council. The committee focused on whether EPA fairly and soundly evaluated the evidence from diverse sources (in vitro, animal, mode-of-action, and epidemiologic studies) and on whether the SOTS evaluation provided a robust, objective, and repeatable analysis. The committee reviewed how clearly EPA describes how its assessment was performed, whether consistent methods and criteria were applied in the analysis of different evidence streams, and whether appropriate methods were applied in evaluating the evidence. Thus, the focus was on the process of the SOTS evaluation and documentation that resulted. The committee organized its evaluation according to the three central scientific questions that EPA sought to answer. Question 1 is considered first, and Questions 2 and 3 are addressed together later.

STATE-OF-THE-SCIENCE APPROACH TO ANSWERING QUESTION 1

To answer the question of whether NMDR curves exist for chemicals and, if so, under what conditions, EPA focused on evidence of NMDR relationships in studies of chemicals that have effects on the estrogen, androgen, and thyroid hormone pathways. EPA's SOTS evaluation made a distinction between NMDR relationships and the related issue of low-dose effects and successfully avoided conflating the two concepts. It presented evidence from in vitro studies and in vivo studies of aquatic species and mammalian models. It concluded that exposure to endocrine disruptors 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.

Because of time constraints, independent groups wrote the sections on the estrogen, androgen, and thyroid hormone pathways. No plan was established in

advance for the writing groups to follow. Rather, the groups determined independently how to perform their analyses. The committee sought to determine how the groups approached their evaluations by looking for documentation of the literature-search strategies, criteria for selecting studies, methods for assessing study quality, presentation of the evidence, and methods used for synthesizing the evidence. Documentation of some of these elements was difficult to find and in most cases had to be inferred. The exception was the section on the mammalian evidence of NMDR curves in studies of thyroid disruptors, which provided enough description of how the literature was searched and filtered, of the modes of action (MOAs) that were considered, and of how the findings were synthesized for the committee to understand the process that was used. EPA acknowledged that such a process was not followed by the groups evaluating the data on the estrogen- and androgen-hormone pathways and that an expert-driven approach was used instead. Such an approach might be appropriate as an internal scoping exercise for the agency, but a higher standard of documentation and analysis is needed to provide the foundational support necessary to make decisions about the agency's toxicity-testing strategies and risk-assessment practices. A lack of transparency and a lack of consistency were identified as two critical limitations of EPA's evaluation. EPA has acknowledged such limitations and has indicated that it plans to conduct more systematic literature searches, data extraction, and evaluations of the evidence on NMDR curves. However, the results of these endeavors were not available to the committee, so findings and recommendations are restricted to what is presented in the SOTS evaluation.

Recommendation: An analytic plan should be developed and applied consistently to the evidence on the three hormone pathways. Important elements of the plan include predefining and documenting the literature-search strategies and their results, establishing criteria for selecting studies for analysis, establishing criteria for determining study quality, using templates for presenting evidence consistently in tabular and graphic form, and documenting approaches to integration of evidence. Guidance on these elements is provided below.

Scoping and Framing Questions

The committee supports EPA's approach of posing questions as a means of focusing the evaluation of the evidence on NMDR curves. The three central scientific questions are critical and are framed broadly. The first question is whether endocrine disruptors have NMDR curves and, if so, under what conditions. Identifying the "conditions" under which NMDR curves occur requires analysis of the different types of evidence that are available. However, EPA restricted its analysis to in vitro studies, studies of aquatic species, and studies of mammalian models; epidemiologic and other types of human studies were excluded from consideration. That exclusion raises a problem in that the SOTS

evaluation will ultimately inform decisions about *human health* risk-assessment practices. In addition, for the estrogen and androgen hormone pathways, it seems that evidence-gathering encompassed only a subset of MOAs by which chemicals and endogenous hormones can act in the body. That restricted the array of chemicals considered to be disruptive of the estrogen and androgen pathways and limited the ability to detect conditions under which NMDR curves might occur as a result of different modes of action.

Recommendation: If EPA wishes to determine the conditions under which NMDR curves occur, it should expand its evaluation to include evidence from epidemiologic and clinical studies and include chemicals that have a variety of potential MOAs for the different hormone pathways. If such a broad analysis is not feasible in light of the agency's immediate needs, consideration should be given to narrowing the questions and their answers.

Methods of Analysis

Systematic approaches to evaluating the scientific literature have been recommended by other National Research Council committees and are being developed by government agencies and incorporated into environmental health assessments. Emphasis is placed on specifying and documenting the methods that will be used to answer a study question. Such methods include determining the literature-search strategy, defining criteria to select studies for analysis, evaluating study quality, presenting data, and integrating evidence. It is noteworthy that EPA's SOTS evaluation of NMDR curves related to chemicals that affect the thyroid hormone pathway contained elements of that type of structured review, but the sections on the estrogen and androgen pathways did not. EPA has indicated that it has since started incorporating elements of systematic review into its literature evaluation in connection with endocrine disruptors.

An issue that should be further explored by EPA is whether the available data lend themselves to answering questions about NMDR curves in a definitive manner. That will require establishing the study-design criteria that would be necessary to characterize whether an NMDR relationship exists and identifying studies that meet the criteria. Conventional toxicity tests have design limitations that can make it difficult for them to detect or characterize NMDR relationships, such as the number of doses, the number of test animals, and the variety of end points considered. Statistical methods are available for conducting post hoc analyses of the data and for combining evidence from multiple studies, including meta-analytic approaches. Such methods could be adapted to evaluating the evidence on NMDR curves. However, that would need further research and development before implementation, inasmuch as current methods have been designed to examine single measures and outcomes and do not have the capacity to explore more complex relationships.

Recommendations: EPA should consistently use a more systematic approach to evaluating the literature on NMDR curves for all three hormone pathways. Guidance for such approaches is available from clinical guidelines, other National Research Council reports, and guidelines being developed in other government agencies. In developing an approach for the literature on NMDR relationships, special consideration should be given to the following:

- The methodologic features that would be necessary for a study to be able to detect an NMDR relationship should be identified. Ideally, multiple dose groups would be spaced across a defined exposure domain, including doses below those typically tested. Statistical design, biologic plausibility, and replicability should be factored into interpreting and weighing the evidence from such studies.
- Study exclusion and inclusion criteria should be established. Although statistical significance is an important consideration, it should not be an absolute criterion for including or excluding studies, inasmuch as standard toxicity-testing strategies generally do not have sufficient sensitivity and statistical power to detect NMDR curves.
- Study quality criteria should be established. Statistical criteria should be given particular attention. It will be important to balance study quality criteria that are based on statistical significance and those based on biologic plausibility.
- Secondary analyses of other studies may be necessary. Current methods for performing post hoc analysis of data and for combining evidence from multiple studies might be adapted for such purposes but would require research and development before implementation. EPA should consider soliciting input from the biostatistics community on the best methods to pursue in the long term and on what measures to take to complete the SOTS evaluation.
- Justifying the use of definitions that are not consistent with those used by EPA programs or that are controversial, such as definitions of low-dose effect, resilience, and adverse effect.

STATE-OF-THE SCIENCE APPROACH TO ANSWERING QUESTIONS 2 AND 3

EPA's SOTS evaluation concluded that current testing strategies include assays that detect chemicals that interfere with the estrogen, androgen, and thyroid signaling pathways. Assertions were made that the evidence is insufficient to show that NMDR curves for adverse effects occur below particular thresholds—NOAELs or benchmark doses (BMDs)—derived from current testing strategies, that current testing assays are sensitive in detecting chemicals that interfere with the hormone pathways, and that NMDR curves occur mainly at high doses. However, it is acknowledged that traditional toxicity testing may not be sufficient for defining the shapes of dose–response curves. These conclusions appear to be based on the expert opinions of the authors of the document. The

present committee recognizes that expert judgment is integral to scientific analyses, but EPA's evaluation provided insufficient documentation of the analyses that led to the conclusions.

The "testing strategies" that were under consideration were not specified, so it was unclear to the committee whether they included EPA's toxicity-testing studies, testing associated with the Endocrine Disruptor Screening Program, testing as part of the ToxCast program, or other types of testing that are important for framing MOA or WOE analysis. Each of those categories of studies has different goals, so EPA's conclusions about the adequacy of toxicity testing should be rooted in specific reference to and demonstration of its own testing strategies in a systematic manner for the estrogen, androgen, and thyroid pathways.

With respect to the question of whether NMDR curves provide information that would alter EPA's current WOE conclusions and risk-assessment determinations, EPA concludes that NMDR curves can have both qualitative and quantitative effects and that current risk-assessment practices will consider the evidence appropriately. However, no analyses of how NMDR curves might affect considerations of hazard identification and dose–response relationships or of how current practices are equipped to handle evidence of NMDR relationships are provided in support of those conclusions. Thus, whether EPA's conclusions are supported by the evidence presented in the SOTS evaluation could not be determined by the committee.

Recommendations:

- EPA should specify the toxicity-testing strategies under consideration and evaluate them transparently for their ability to detect NMDR relationships mediated by the estrogen, androgen, and thyroid hormone pathways. The tests should be evaluated in terms of the design criteria that would be necessary to detect such relationships and the adverse effects of interest.
- The concept of "resilience" or adaptation is controversial and not adequately addressed by EPA in the SOTS evaluation. EPA should distinguish between end points that are adverse and ones that are adaptive. Because effects that are adaptive in some people are adverse in others, consideration should be given to potential windows of susceptibility (for example, during fetal development), sensitive populations (for example, those with pre-existing health conditions), and other factors (such as multiple chemical exposures) in making these distinctions.
- EPA's evaluation should indicate how NMDR relationships for estrogen, androgen, and thyroid pathways would be addressed under the agency's current risk-assessment guidelines and practices, including statistical considerations, uncertainty analyses, life-stage or susceptibility issues, and MOAs.

1

Introduction

Potential health effects of endocrine-disrupting chemicals have been an environmental health concern for decades because of their ability to interfere with normal hormone function of humans and wildlife populations. Endocrine-disrupting chemicals are exogenous substances or mixtures that alter function(s) of the endocrine system and so may cause adverse health effects in an intact organism or its progeny or subpopulations (EPA 2013). Endocrine disruptors are heterogeneous and include synthetic chemicals used as industrial solvents and their byproducts (such as polychlorinated biphenyls and dioxins), plastics (such as bisphenol A), plasticizers (such as phthalates), pesticides (such as chlorpyrifos and DDT), fungicides (such as vinclozolin), pharmaceutical agents (such as diethylstilbestrol), and natural chemicals found in human and animal food (such as phytoestrogens) (e.g., Diamanti-Kandarakis et al. 2009).

Because endocrine systems exhibit tissue-, cell-, and receptor-specific actions through the life course of organisms, the mechanisms by which endocrine disruptors may interfere with hormone function are complex and challenging to decipher (Zoeller et al. 2012). Evolving epidemiologic, toxicologic, mechanistic, and biomonitoring research on endocrine disruptors has invigorated the discussion with respect to their effects, the mechanisms by which the effects occur, and the interpretation of the data for science-policy implications. Concerns also have been raised about nonmonotonic dose–response (NMDR) relationships reported in in vitro, in vivo, and epidemiologic studies because conventional assumptions about chemical toxicity may not be applicable to chemicals that have such dose–response curves. NMDR curves are characterized by a change in the sign of their slope. They may be U-shaped, inverted U-shaped, or more complex (see Figure 1-1). NMDR curves could have implications for regulatory toxicity testing and risk assessment, which typically extrapolate from higher doses in animals to lower exposures in the environment for testing the safety of chemicals and thus assume that the slope of the dose–response curve does not change sign, that is, is monotonic. Controversy regarding this issue led the US Environmental Protection Agency (EPA) to review the evidence on NMDR curves and to request that the National Research Council conduct an independent scientific review of the draft of *State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Path-*

ways and EPA Testing and Assessment Procedures (referred to here as the SOTS evaluation). In response to EPA's request, the National Research Council convened the Committee to Review EPA's Draft State of the Science Paper on Nonmonotonic Dose Response, which prepared the present report.

BACKGROUND AND HISTORY OF ENDOCRINE-DISRUPTOR RESEARCH

EPA has a long history of involvement with endocrine-disruptor research. This section highlights a number of issues that have focused scrutiny on NMDR relationships and that have led to the development and review of EPA's draft SOTS evaluation.

The Food Quality Protection Act of 1996 mandated that EPA "develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect" (EPA 2011). As a result, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) was established to advise EPA on methods of screening and testing individual chemicals for endocrine-disruptor activity. EDSTAC (1998) proposed a two-tier system in which an initial battery of relatively short-term in vitro and in vivo assays to screen chemicals for potential endocrine-disruptor activity (activity that may interfere with estrogen, androgen, and thyroid hormones) would be used (Tier 1), followed by a second set of more refined tests (Tier 2). Since then, EPA's Endocrine Disruptor Screening Program has further developed the Tier 1 assays and engaged in a validation process intended to ensure that the Tier 1 data would be reliable and reproducible among laboratories. More recently, EPA has embarked on the use of high-throughput screening in vitro assays that would be faster, would be more efficient, and would make it possible to evaluate a greater number of chemicals (Rotroff et al. 2013).

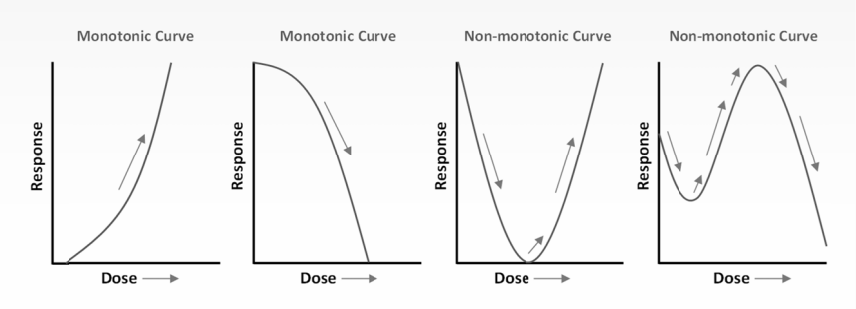

FIGURE 1-1 Examples of monotonic and nonmonotonic dose–response curves. Source: EPA 2013 (Adapted from Fagin 2012).

In 2000, EPA and the National Institute of Environmental Health Sciences (NIEHS) held a workshop on low doses of endocrine-disrupting chemicals (EDCs) to provide an independent peer review of the available data, including a statistical reanalysis of studies that suggested the existence of NMDR relationships for EDCs. Among the findings from the workshop were that there was sufficient evidence of low-dose reproductive and developmental effects and that some estrogenic compounds exhibit NMDR curves (NTP 2001).

In 2011, EPA held a workshop on low-dose effects of EDCs—with participation by EPA scientists of the Office of Chemical Safety and Pollution Prevention (OCSPP), the Office of Water, and the Office of Research and Development (ORD)—to review the state of the science and discuss its potential implications. Workshop participants emphasized the need for a comprehensive scientific review given the rapid advances in this field (Schoeny 2013).

In 2012, a number of scientific publications on NMDR curves and EDCs received a great deal of attention from the scientific, nongovernment, and industry communities. Vandenberg et al. (2012) reviewed the extensive literature on NMDR relationships, and concluded that "fundamental changes in chemical testing and safety determination are needed to protect human health." In an editorial in *Environmental Health Perspectives*, the director of NIEHS cited that review and indicated that future research should be directed at answering the question of "which dose-response shapes should be expected from specific environmental chemicals and under what specific circumstances" (Birnbaum 2012). A commentary by Rhomberg and Goodman (2012) on the Vandenberg et al. review was critical of the approach used to evaluate the literature and draw conclusions. The Endocrine Society published a position statement in 2012 that provided guidelines for applying fundamental principles of endocrinology to identify EDCs and to assess their potential risks (Zoeller et al. 2012). Those and other publications and the discourse surrounding them highlight the critical need for using a transparent, well-defined, and clearly articulated strategy in seeking to assess the effect of NMDR relationships on toxicity-testing strategies and how they affect human health and environmental risk assessments.

Several European and international organizations have evaluated whether the evidence on EDCs and NMDR curves requires changes in testing strategies. For example, the Centre on Endocrine Disrupters of the Danish National Food Institute evaluated the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) testing guidelines (Hass et al. 2013) and included critiques of the Vandenberg et al. (2012) paper. It concluded that "the current information requirements in REACH are not designed for the identification of endocrine disrupters, although certain endpoints and assays may give some indication of endocrine disrupting effects. It is, however, evident that important endpoints needed for the detection of [endocrine-disrupter] effects are not included." The UN Environment Programme and World Health Organization (UNEP/WHO 2013) published *State of the Science of Endocrine Disrupting Chemicals—2012*. The report noted the need for revising current testing strategies because of questions, raised by the evidence on EDCs, about the adequacy

of typical testing exposure paradigms and the array of end points considered. A scientific committee of the European Food Safety Authority (EFSA) conducted a review of the scientific criteria for identifying EDCs and of methods for assessing effects mediated by such chemicals (EFSA 2013). The EFSA committee concluded that "a reasonably complete suite of standardized assays for testing the effects of [EDCs] is (or will soon be) available for the oestrogenic, androgenic, thyroid and steroidogenic modalities in mammals and fish, with fewer tests for birds and amphibians." It also stated that it "cannot conclude whether the test methods are adequate to fully define dose response relationships. However, the available information is equally insufficient to conclude that current dose response analysis in regulatory (eco)toxicology should be modified on a routine basis."

EPA's ORD began work on the draft SOTS evaluation in 2012 in response to a request from OCSPP. The agency convened a cross-program, cross-agency working group that engaged other federal partners (for example, the Food and Drug Administration and the National Institute of Child Health and Human Development) to provide input for the evaluation. The working group adopted consensus definitions of four key terms for the SOTS evaluation: *NMDRs, EDC, low-dose effect,* and *adverse effect*. The definitions were:

- NMDRs: "measured biological effects with dose response curves that contain a point of inflection where the slope of the curve changes sign at one or more points within the tested range" (EPA 2013).
- EDC: "an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations" (adopted from WHO 2002).
- Low-dose effect: "a biological change occurring in the range of typical human exposures or at doses lower than those typically used in standard testing protocols" (adopted from NTP 2001).
- Adverse effect: "a measured endpoint that displays a change in morphology, physiology, growth, development, reproduction, or life span of a cell or organism, system, or population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences" (adopted from Keller et al. 2012).

Four subgroups worked independently on the sections of the evaluation that dealt with estrogen, androgen, and thyroid (in human and nonhuman) hormone pathways. The groups used various approaches for searching and reviewing the literature (details of these approaches are in Chapter 2). They focused explicitly on EDCs that act through the estrogen, androgen, and thyroid hormone systems. Because of the limitations of the literature, the focus was limited to single chemical exposures. The draft SOTS evaluation states that the report is not intended to be a policy document or to reflect testing guidance but rather as a resource for informing decision-making. EPA intends to revise the document in response to recommendations made in the present report and after considering

public comments. To the committee's knowledge, this is one of the few attempts to evaluate specifically the implications of the evidence on NMDR curves for testing strategies and risk-assessment practices, and the agency is commended for undertaking such a challenging task.

THE COMMITTEE'S TASK AND APPROACH

Given the complex nature of NMDR curves exhibited by EDCs and their possible importance in interpreting toxicity testing and risk assessment, EPA asked the National Research Council to review the draft SOTS evaluation. In response to the request, the Research Council convened the Committee to Review EPA's Draft State of the Science Paper on Nonmonotonic Dose Response. The committee was charged with reviewing EPA's draft SOTS evaluation, commenting on EPA's analysis of the existence of NMDR curves, and considering the implications of NMDR curves for chemical-testing and risk-assessment practices. The complete statement of task is presented in Box 1-1.

BOX 1-1 Statement of Task

An ad hoc committee will conduct a scientific review of EPA's draft paper, *State of the Science on Nonmonotonic Dose Response*. Specifically the committee will review and provide a brief report on the following:

- EPA's analysis of the potential existence of nonmonotonic dose-response (NMDR) curves for chemicals. Has EPA fairly and soundly evaluated the weight of evidence and has it reached conclusions supported by the available studies?
- EPA's evaluation of the studies and expert opinion (including the completeness of the database) used to assess whether current chemical testing strategies capture *adverse effects* potentially represented by NMDR curves. To what extent do the available studies capture adverse effects?
- EPA's scientific rationale used to evaluate whether the state of the science influences EPA's weight of evidence conclusions and the implications for risk assessment.

For the issues addressed in the bullets above, the committee will consider:

  - Is EPA's *State of the Science* document scientifically sound and of high quality?
  - Has EPA selected studies of suitable breadth, relevance, and quality?
  - Has EPA fairly and soundly evaluated and integrated the weight of evidence from the diversity of studies (epidemiological, mode-of-action, animal testing)?
  - Are the assumptions valid and reasonable?
  - Are the conclusions valid and supported on the basis of EPA's assessment and the literature?
  - Are there potential limitations or data gaps that would substantially impact the conclusions?

To address its task, the committee held three meetings. At the first meeting, EPA officials provided an overview of the development process that led to the draft SOTS evaluation, and an open-microphone session was held to hear the views of interested stakeholders on the evaluation. The second and third meetings were used to draft this report. Throughout the course of its work, the committee considered input from interested stakeholders and additional input that it requested from EPA. As specified in its charge, the committee focused on determining whether the draft SOTS evaluation provides a credible overview of the evidence and adequately documents the methods used to analyze the evidence and on determining whether conclusions were appropriately drawn and justified. The committee did not conduct an independent literature review to draw its own conclusions about the evidence on NMDR curves.

ORGANIZATION OF THIS REPORT

The committee's review of the SOTS evaluation is presented in the following two chapters. Chapter 2 reviews the approach and methods that EPA used for its literature search, study selection, study evaluation, and synthesis and integration of the evidence on NMDR curves. Chapter 3 reviews how EPA assessed the implications of NMDR curves for the agency's testing strategies, its weight-of-evidence conclusions, and its risk-assessment determinations.

REFERENCES

Birnbaum, L. 2012. Environmental chemicals: Evaluating low-dose effects. Environ. Health Perspect. 120(4):A143-A144.

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

EDSTAC (Endocrine Disruptor Screening and Testing Advisorty Committee). 1998. Final Report [online]. Available: http://www.epa.gov/endo/pubs/edspoverview/finalrpt.htm [accessed Sept. 2, 2013].

EFSA (European Food Safety Authority). 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.

EPA (U.S. Environmental Protection Agency). 2011. Endocrine Disruptor Screening Program. Endocrine Primer [online]. Available: http://www.epa.gov/endo/pubs/edspoverview/primer.htm [accessed Aug. 30, 2013].

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: http://epa.gov/ncct/download_files/edr/NMDR.pdf [accessed Aug. 12, 2013].

Fagin, D. 2012. Toxicology: The learning curve. Nature 490(7421):462-465.

Hass, U., S. Christiansen, M. Axelstad, K. Dreisig Sorensen, and J. Boberg. 2013. Input for the REACH-review in 2013 on Endocrine Disrupters. Final report. Center on Endocrine Disrupters, Division of Toxicology and Risk Assessment, National Food Institute, Technical University of Denmark. March 21, 2013 [online]. Available: http://www.mst.dk/NR/rdonlyres/B865F94A-54C0-43DF-AB0F-F07A593E1FD0/0/ReachreviewrapportFINAL21March.pdf [accessed July 15, 2013].

Keller, D.A., D.R. Juberg, N. Catlin, W.H. Farland, F.G. Hess, D.C. Wolf, and N.G. Doerrer. 2012. Identification and characterization of adverse effects in the 21st century toxicology. Toxicol. Sci. 126(2):291-297.

NTP (National Toxicology Program). 2001. National Toxicology Program's Report of the Endocrine Disruptors Low Dose Peer Review. National Toxicology Program, U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC [online]. Available: http://www.epa.gov/endo/pubs/edmvs/lowdosepeerfinalrpt.pdf [accessed Aug. 30, 2013].

Rhomberg, L.R., and J.E. Goodman. 2012. Low-dose effects and nonmonotonic dose-responses of endocrine disrupting chemicals: Has the case been made? Regul. Toxicol. Pharmacol. 64(1):130-133.

Rotroff, D.M., D.J. Dix, K.A. Houck, T.B. Knudsen, M.T. Martin, K.W. McLaurin, D.M. Reif, K.M. Crofton, A.V. Singh, M. Xia, R. Huang, and R.S. Judson. 2013. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ. Health Perspect. 121(1):7-14.

Schoeny, R. 2013. State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures. Presentation at the First Meeting on Review of EPA's Draft Paper State of the Science on Nonomonotonic Dose Response, July 23, 2013, Washington, DC.

UNEP/WHO (United Nations Environment Program/World Health Organization). 2013. State of the Science of Endocrine Disrupting Chemicals-2012, A. Bergman, J.J. Heindel, S. Jobling, K.A. Kidd, and R.T. Zoeller, eds. World Health Organization [online]. Available: http://www.who.int/ceh/publications/endocrine/en/ [accessed Aug. 30, 2013].

Vandenberg, L.N., T. Colborn, T.B. Hayes, J.J. Heindel, D.R. Jacobs, D.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.

WHO (World Health Organization). 2002. Global Assessment of the State-of-the-Science of Endocrine Disruptors. International Programme on Chemical Safety. World Health Organization [online]. Available: http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/ [accessed Aug. 30, 2013].

Zoeller, R.T., T.R. Brown, L.L. Doan, A.C. Gore, N.E. Skakkebaek, A.M. Soto, T.J. Woodruff, and F.S. Vom Saal. 2012. Endocrine-disrupting chemicals and public health protection: A statement of principles from The Endocrine Society. Endocrinology 153(9):4097-4110.

2

Review of the Environmental Protection Agency's Methods for Evaluating Evidence on Nonmonotonic Dose–Response Relationships

INTRODUCTION

This chapter reviews the methods used by the US Environmental Protection Agency (EPA) to identify, evaluate, and synthesize the scientific evidence on nonmonotonic dose response (NMDR) curves in its draft State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures (the SOTS evaluation; EPA 2013a). To address its first task, to determine whether EPA "fairly and soundly evaluated the weight of evidence" on NMDR curves, the committee first identified the key design elements of a thorough, systematic, and transparent evaluation and synthesis of environmental health data (see Box 2-1).

Those elements are derived from accepted approaches to literature-based evidence synthesis in the clinical sciences, particularly methods for systematic review in clinical medicine (Guyatt et al. 2011; Higgins and Green 2011). The National Research Council (NRC 2011, 2013) has recommended that EPA use similar approaches to support and improve its toxicologic assessments in support of its Integrated Risk Information System program. The methods in clinical medicine are not directly transferable, because environmental health assessments include evidence from multiple lines of research (in vitro, animal, and human studies) whereas clinical-medicine evaluations are based exclusively on studies of humans (Woodruff and Sutton 2011; Birnbaum et al. 2013). However, modified methods have been proposed for evaluating environmental health evidence (e.g., Woodruff and Sutton 2011; Kushman et al. 2013), and efforts are under way in EPA (2013b) and the National Toxicology Program (NTP) Office of Health Assessment and Translation (NTP 2013) to incorporate systematic approaches into their toxicologic assessments. Although the committee recognizes that such a systematic approach has not yet been formally established,

BOX 2-1 Design Elements of a Systematic Review

- Define study question
- Specify methods for collecting and evaluating evidence
  - Literature-search strategy
  - Study inclusion and exclusion criteria
  - Methods for evaluating study quality
  - Data presentation
  - Methods for analyzing and synthesizing evidence

EPA has received enough guidance and recommendations from other National Research Council reports (NRC 2009, 2011) to have considered the use of more consistent and transparent approaches similar to the design elements shown in Box 2-1 to develop the SOTS evaluation.

The committee thus used the following criteria to evaluate how each of the elements in Box 2-1 was addressed in the SOTS evaluation:

- Clarity: Is the SOTS evaluation clear in its description of how each element was addressed?
- Consistency: Is the SOTS evaluation consistent among topics in its application of methods and criteria?
- Appropriate methods: Were the SOTS evaluation's approaches to evaluating evidence appropriate?

Those criteria were used to examine the assessments that were conducted both within and among the three hormone pathways considered (estrogen, androgen, and thyroid) and to evaluate the different streams of evidence (in vitro studies and in vivo studies of aquatic species and animal models).

EVALUATION OF THE ENVIRONMENTAL PROTECTION AGENCY'S APPROACH

EPA's strategy for developing the SOTS evaluation was to pose three central scientific questions about NMDR curves with respect to the estrogen, androgen, and thyroid hormone pathways. The agency was faced with the difficult challenge of comprehensively identifying, evaluating, and summarizing the large volume of information required to address those questions. The foreword to the SOTS evaluation indicates that because of time constraints the agency used an expert-driven approach to conduct the evaluation, which involved having different groups evaluate the evidence on the three hormone pathways separately. However, EPA did not establish an analysis plan in advance for the writing groups to follow. Rather, the groups determined independently how to

perform their analyses. That led to the lack of an overall plan for the approach, differences in the degree of documentation provided about the literature search and selection process used by the groups, differences in criteria that were used for study selection, and unclear documentation of how study quality was evaluated and of how conclusions were drawn. The lack of transparency and the inconsistencies raise questions about the quality of the approaches used.

Lack of Transparency. There was a lack of transparency at key steps of the SOTS evaluation. Prominent examples include the lack of documentation of the literature-search methods used for the estrogen and androgen sections, the lack of explicit criteria for evaluating study quality, the lack of clarity of the weight-of-evidence (WOE) methods, and the inadequate explanation of the process used to identify which studies carry the most weight (and why). Furthermore, the methods of data synthesis and weighing of evidence to identify conditions under which NMDR curves occur were not presented transparently. Thus, it was unclear how the authors concluded that NMDR curves were found more often in in vitro studies, at high doses, and for exposures of short duration.

Inconsistency. Inconsistencies were found in the methods used to identify studies for consideration; in the study inclusion and exclusion criteria and their application; in the criteria for evaluating study quality; in how data were presented, weighed, and analyzed; and in how the key data for each section were summarized. Those inconsistencies resulted largely from not having established a protocol for the writing groups to follow or, if a protocol was established, from failure to adhere to it rigorously for each of the three different hormone modalities. In the absence of a clear framework and its consistent application, the separate writing groups inevitably used different methods of review; this calls into question whether a more systematic approach would have led to the same conclusions.

The sections below review specific aspects of the SOTS evaluation that led the committee to draw those overarching conclusions.

Protocol

EPA decided to use independent groups to draft the SOTS evaluation, but there was no protocol for performing the evaluations to ensure that the different groups followed the same methods to reach conclusions. Instead, each group was allowed to perform its analysis on the basis of expert judgment. Documentation of the methods used by each group was difficult to find and in most cases did not appear to be provided.

Stipulating the methods ahead of an assessment and then applying them consistently in the various sections is a means of reducing author bias and of providing transparency. Several National Research Council reports have highlighted the importance of planning and scoping. For example, *Science and Decisions: Advancing Risk Assessment* noted that "increased emphasis on planning

and scoping and on problem formulation has been shown to lead to risk assessments that are more useful and better accepted by decision-makers" (NRC 2009, p. 6). Chapter 7 of the 2011 report on formaldehyde called for EPA to "ensure standardization of review and evaluation approaches among contributors and teams of contributors; for example, include standard approaches for reviews of various types of studies to ensure uniformity" (NRC 2011, p. 164).

Study Questions

Defining the study questions determines the structure and scope of any assessment (IOM 2011). The committee found the three central scientific questions of the SOTS evaluation (Box 2-2) to be clear and reasonable. Question 1 is framed broadly and is open-ended with respect to determining the "conditions" under which NMDR curves might occur. The scope of the question suggested to the committee that EPA would evaluate all relevant streams of evidence to answer it. However, the SOTS evaluation was restricted to in vitro studies and in vivo studies of aquatic species and animal models. Because of inadequate resources, evidence from epidemiologic or other types of human studies is not considered beyond reference to reviews conducted by other groups (e.g., Vandenberg et al. 2012; EFSA 2013). Furthermore, the scope of conditions being considered was not specified. It appears that the evidence on the estrogen and androgen hormone pathways was restricted to chemicals that have narrowly defined modes of action (discussed below under "Study-Selection Criteria"); this would limit the mechanistic conditions under which an NMDR curve might occur. There is an important incompatibility between the broad question posed by EPA and the narrow array of data considered to answer it. That is problematic because the answers to questions about the adequacy of EPA's toxicity-testing strategies and risk-assessment practices (Questions 2 and 3) depend critically on the scope of the answer to Question 1.

The SOTS evaluation also uses several definitions that are important in determining the scope of the answers to the questions posed. The committee offers several observations about three key definitions:

BOX 2-2 Three Central Scientific Questions to be Addressed in the SOTS Evaluation (EPA 2013a)

- **Question 1:** Do [NMDRs] exist for chemicals and if so under what conditions do they occur?
- **Question 2:** Do NMDRs capture adverse effects that are not captured using [EPA's] current chemical testing strategies (i.e., false negatives)?
- **Question 3:** Do NMDRs provide key information that would alter EPA's current weight of evidence conclusions and risk assessment determinations, either qualitatively or quantitatively?

- *Low-dose effect* is defined in the SOTS evaluation as "a biological change occurring in the range of typical human exposures or at doses lower than those typically used in standard testing protocols", which is the definition used by NTP (2001). That definition is vague and confuses two concepts—"low dose" and "low effect"—both of which are important in understanding endocrine disruptors and for identification of NMDR curves. Just the low-dose part of the definition is nonspecific: the range of doses reported as "low dose" in the literature and those used by NTP and relevant human exposures can differ by orders of magnitude (Teeguarden and Hanson-Drury 2013). Being clear about what is meant by *low dose* is thus critical to study evaluation and interpretation. The definition of *low effect* clearly depends on the end points being measured, and could be an issue at any dose, including those in the range used by NTP in animal bioassays. Thus, the committee recommends that EPA's definitions be clear and specific to ensure their consistent application throughout the SOTS evaluation.
- *Resilience* is described in the SOTS evaluation as the ability of cells and tissues to adapt to maintain homeostasis. The concept of resilience, or adaptation, in a toxicologic context has varied definitions and is a controversial topic that requires careful consideration, but the overview in the SOTS evaluation (p. 30, Figure 2.2) is brief and insufficiently supported, with only a single reference (Andersen et al. 2005). The figure presented from that reference was adapted in a National Research Council report (NRC 2007) to address an important limitation of the resilience concept: that it might not apply in all cases and situations. The caption of the revised figure states that "when perturbations are sufficiently large or when the host is unable to adapt because of underlying nutritional, genetic, disease, or life-stage status, biologic function is compromised, and this leads to toxicity and disease" (NRC 2007, p. 49). The human population is composed of individuals in various states of disease and adaptation. Adaptation may occur in healthy adults who are exposed to a single chemical; but if there are multiple chemical exposures or exposures occur during critical periods of development, there could be little or no ability to adapt or change (Woodruff et al. 2008). Thus, there is particular concern about how the concept of resilience might have been used to evaluate data resulting from studies in which exposure occurred during development. It is also unclear how consideration of resilience might have affected study selection. The revised SOTS evaluation should either elaborate on this concept and specify how it was used in the analysis or omit it, particularly if it was provided simply for reference but did not affect study selection.
- *Adverse effect* is defined in the SOTS evaluation as "a measured endpoint that displays a change in morphology, physiology, growth, development, reproduction, or life span of a cell or organism, system, or population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences" (adopted from Keller et al. 2012). The SOTS evaluation indicates that this defi-

nition is preferable to the one used by the agency for health-assessment purposes[1] because it is a more "systems biology oriented description of adversity" and allows for consideration of mode of action, toxicity pathways, adverse-outcome pathways, and adaptive capability. Better justification for using the definition is needed, especially because the SOTS evaluation is intended to inform decisions about risk-assessment practices; explicit consideration should be given to the implications of this definition, in contrast with the definition used by EPA risk assessors, for health end-point selection, study selection and weighting, and related decisions.

Literature-Search Strategy

A literature-search strategy is designed to identify the universe of potentially relevant studies once a study question is specified. Information about the strategy and results should be provided in sufficient detail to ensure that each database search is replicable. For example, the databases should be specified, the search terms and strings listed, the dates on which the searches were conducted identified, and a summary of the search results provided. Any examination of the literature that is not thorough and systematic runs a risk of assembling an unrepresentative selection of publications and could lead to erroneous conclusions.

The independent writing groups of the SOTS evaluation conducted their literature searches differently. In most cases, no documentation of the search strategies was provided. An expert-based approach to literature selection and analysis appears to have been used for the estrogen and androgen sections. Section 4.2.1 ("Literature Search and Selection Strategy for Estrogen and Androgen Pathways") does a reasonable job of describing the complexities of the literature but says only in general terms that a "large database of journal articles and other reports were examined" and does not adequately describe how studies were identified and selected. In the androgen section, Table 4.3 presents 29 studies that were used to evaluate the androgen hormone pathway, but the committee could not find any description of a search strategy that led to their selection. Similarly, adequate documentation was not provided in the thyroid section on aquatic species (Section 4.1.4). In contrast, the search strategy is presented with reasonable completeness in the thyroid section on mammalian models (Section 4.2.4.2 and Appendix C). Failure to establish a clear method of literature identification and evaluation for the groups to follow is responsible for the section-to-section variation and is an important shortcoming of the report. EPA is aware of this problem and has indicated that it is monitoring systematic approaches being developed by the National Toxicology Program and EPA's Integrated Risk In-

[1]"A biochemical change, functional impairment, or pathologic lesion that affects the performance of the whole organism, or reduces an organism's ability to respond to an additional environmental challenge" (EPA 2014).

formation System program and will consider them in revisions to the SOTS evaluation (EPA, unpublished material, September 6, 2013).

Study-Selection Criteria

In systematic reviews, clearly defined eligibility criteria are used to determine which studies will be included and which will be excluded from evaluation. Explicit and well-defined criteria are fundamental for a rigorous gathering of a defensible set of data for review (Abrami et al. 1988; Meline 2006), and they provide a guide for the standard of research used to evaluate WOE. Selection criteria should be formulated to identify and include as many informative studies as possible. Several approaches are available to guide study selection. One is to develop a population–exposure–comparator–outcome statement to define each of the elements of the studies. That approach is being modified for application to environmental and toxicologic studies by NTP (2013) and others (e.g., Koustas et al. in press). EPA should consider and adopt an approach that clearly lays out its study-selection criteria.

In the sections below, the committee considers whether the selection criteria in the SOTS evaluation were clearly presented and whether they would ensure that an appropriate set of studies is considered in the analysis.

General Issues with Study-Selection Criteria in the State-of-the-Science Evaluation

EPA's SOTS evaluation states that it did not "attempt to design exclusion/inclusion criteria for studies uncovered in all literature searches; for reasons of resource limitations, this was done primarily for the description of the data on the thyroid hormone pathway" (EPA 2013a, p. 27). Thus, the literature search and analysis for thyroid disruptors (Section 4.2.4.2) describe the study-selection criteria and present a decision tree in an appendix (Figure C.1) to illustrate how the criteria were used to filter studies. Although the criteria were clearly laid out for the thyroid section, the committee questions the appropriateness of some of the criteria used to filter the studies (see section "Selection Criteria for Thyroid-Active Chemicals" below).

The committee attempted to identify some of the informal study-selection and study-quality criteria used in the estrogen and androgen sections. Table 2-1 presents a comparison of the study-selection criteria for the three hormonal pathways. For all three pathways, an "ideal" number of doses evaluated, type of chemicals evaluated, and specific restrictions for each of the sections are specified. As illustrated in the table, the criteria varied among sections (and even within sections for different chemicals), and the rationale for their use was only partially explained. For instance, Section 4.1 ("Aquatic Models") specifies that studies with four doses are preferred but that studies with fewer doses could be

TABLE 2-1 Comparison of Study-Selection Criteria and Study-Quality Evaluations Used in Different Sections of the SOTS Evaluation

		Study-Selection Criteria			
Section of SOTS Evaluation	SOTS Pages	Number of Dose Groups	Chemicals and Modes of Action Considered	Restrictions and Filters	Study Quality Criteria
3. In vitro studies[a]	43-44	Unable to determine	Unable to determine	Unable to determine	Unable to determine
4.1.1 Aquatic models (HPG axis)	56-57	A. ≥4 treatment groups B. <4 groups (included for chemicals or pathways where A not available)	A. Starting list included 28 "model chemicals" listed in Table 4.1 and was reduced to 11 "best-studied" and "illustrative" examples B. Chemicals with adequately described MOA on the HPG axis (e.g., atrazine excluded because of insufficiently characterized MOA); no explanation of "adequately described"	Restrictions: A. "Restricted largely to" fish species B. Time windows B.1. Full life cycles B.2. When B.1 not available, "longer-term experiments during portions of the life-cycle expected to be sensitive to endocrine-active chemicals"	Unable to determine
4.1.1 Aquatic models (Thyroid)	58	Unable to determine whether criteria used for aquatic models (HPG axis) were applied to this set of studies	Chemicals with effects on either: A. Sodium-iodide symporter (NIS) B. Thyroid peroxidase (TPO)	Restrictions: A. Fish and amphibians were included B. Time windows: unclear whether restrictions in point B above for aquatic models (HPG axis) were applied to this set of studies	Unable to determine
4.2.1 Mammalian models (estrogen and androgen pathways)	84	A. ≥6 groups OR B. ≥4 groups (3 + 1 control), but large range of exposure ("large range" not defined)	No exclusion	Restrictions: A. Oral exposure B. Subcutaneous included if genomic study or if only few or no oral studies for a specific MOA	Unable to determine

(Continued)

TABLE 2-1 Continued

Section of SOTS Evaluation	SOTS Pages	Study-Selection Criteria			
		Number of Dose Groups	Chemicals and Modes of Action Considered	Restrictions and Filters	Study Quality Criteria
4.2.4.2 Mammalian models (thyroid)	121-122	≥3 doses + control	Only one chemical for each of three MOAs described in depth: A. PTU (TPO inhibition) B. Perchlorate (NIS inhibition) C. PHAHs (u-regulation of thyroid hormone metabolism induced by nuclear receptor activation) No exclusions for the broader mammalian thyroid literature	Filters: 1. Minimum of three doses + control Evidence of statistically significant NMDR relationship 2. Absence of observations at lower doses in the study that would have been used to determine the LOEL or LOAEL 3. (a) Absence of other published reports on the chemical in which effects were observed at low levels; (b) absence of other published reports for effects on other end points that would have been used to determine LOEL or NOEL below the doses identified as having an NMDR relationship; (c) absence of study-quality concerns or statistical-power issues that weakened confidence in NMDR observation	Unable to determine

[a]Studies identified by Vandenberg et al. (2012). Description of search strategy and study selection were not provided in the Vandenberg et al. paper.

Abbreviations: HPG, hypothalamic-pituitary-gonadal; LOAEL, lowest observed-adverse-effect level ; LOEL, lowest observed-effect level; MOA, mode of action; NIS, sodium-iodide symporter; NMDR, nonmonotonic dose response; NOEL, no-observed-effect-level; PHAHs, poly-halogenated aromatic hydrocarbons; PTU, propylthiouracil; TPO, thyroperoxidase.

used to include additional chemicals and pathways in the analysis. In contrast, Section 4.2.1 ("Literature Search and Selection Strategy for E and A Pathways"), a minimum of six doses is specified, but in vivo studies that tested fewer doses were allowed for inclusion provided that the studies tested a "broad dose-range" so that studies cited by others as displaying NMDR curves could be included. However, no definition of a broad dose range was provided. Most of the sections restrict the evaluation to subsets of the studies available on the basis of species, route of exposure, or exposure levels. The restrictions do not appear to be consistently applied among sections.

In determining the set of chemicals to evaluate for each of the hormonal pathways, it appeared that the SOTS evaluation was not consistently comprehensive in its consideration of modes of action (MOAs). The section on the thyroid pathway provided the clearest description of the MOAs considered. Known mechanisms of thyroid disruption were described (pp. 36 and 119–120 and Figure 2.5), and the search strategy documents that potential NMDR curves with both genomic and nongenomic actions were considered. The thyroid section also included chemicals that have effects both on and outside the hypothalamic–pituitary–thyroid axis. Thus, this section was appropriately inclusive in considering potential MOAs.

In contrast, the scope of the MOAs considered for the estrogen and androgen pathways was unclear. On the basis of the data presented in the two sections, the evaluations appear to focus on studies in which nuclear receptor-mediated activity was observed or presumed. For example, although it is acknowledged in the text that estrogen can act on the cell via multiple mechanisms, only the nuclear receptor-mediated activity is depicted in Figure 2.6, and it appears that this is how the "estrogenic" MOA is primarily defined in the SOTS document. The committee found this to be too limited an interpretation of estrogen activity. For example, effects of estrogen action via the newly identified receptor GPR30 are dismissed in the SOTS evaluation as "difficult to evaluate" because the "responses are often of very low magnitude, and, although statistically significant, have little biological validation" (EPA 2013a, p. 49). Although investigation of estrogenic activity through GPR30 signaling and the physiologic outcomes associated with it continue, this MOA and similar rapid-signaling pathways initiated by estrogen are proving to be biologically significant (Filardo and Thomas 2012). Excluding this MOA not only is inappropriately dismissive but demonstrates how adoption of a narrowly conceived MOA constrains the identification of "under what conditions do [NMDR curves] occur" (part of Question 1). Not understanding the biologic significance of GRP30 signaling is an important and reasonable consideration in addressing Questions 2 and 3 but is too limiting for addressing Question 1. Also missing is discussion of whether studies reporting epigenetic modifications were included or excluded from consideration. Numerous studies have demonstrated that endocrine disruptors—such as bisphenol A, phthalates, vinclozolin, methoxychlor, and dioxins—can produce epigenetic modifications associated with altered behavioral, reproductive, and other neuroendocrine end points (Dolinoy et al. 2007; Prins et al.

2008; Wolstenholme et al. 2011; Guerrero-Bosagna et al. 2012; Tang et al. 2012; Kundakovic et al. 2013; Manikkam et al. 2013; Somm et al. 2013). It is biologically plausible, and there is precedent for steroid hormones to act through MOAs other than steroid receptor mediation.

The discussion of MOAs relevant to the androgen pathway (p. 36) focuses only on events surrounding androgen-receptor activation (agonism and antagonism) and touches on the effects of disrupting androgen-converting enzymes. However, MOAs and adverse-outcome pathways important in suppressing androgen synthesis (relevant for phthalates) are not discussed although they appear to be considered in subsequent sections analyzing the literature on NMDR curves (pp. 82 ff). In general, the selection of studies for androgenic activity appears to consider all relevant modalities, but the study-selection description lacks the transparency needed to verify that that is the case. Furthermore, in describing the selection strategy for the estrogen and androgen pathways, the SOTS evaluation states that a large database of journal articles and other reports was examined (p. 83), but the database is not identified or described.

Selection Criteria for Thyroid-Active Chemicals

The thyroid section is the only section that explicitly defines the criteria that were used for study selection, which includes four filters, as shown in Table 2-1. Generally, systematic-review methods used in clinical medicine (Higgins and Green 2011; IOM 2011) and those being developed for environmental health assessments (NTP 2013) recommend that selection criteria be chosen on the basis of relevance to the question being evaluated rather than only to study-quality issues. Study quality is important and should be evaluated, but after the studies have been selected for analysis. Study inclusion and exclusion criteria that are too strict may lead to the omission of studies that can provide useful information even if they are limited in some way. Accordingly, the committee disagrees with the use of criteria that would exclude studies because of issues related to statistical power (which are encompassed by Filters 2-4). Using statistical significance as an absolute criterion for selecting studies with respect to NMDR curves is not recommended, because it can be influenced by several factors that should be explored before a determination of how informative a study would be in addressing the question (see Chapter 3 for further discussion of this issue). For example, having a small number of data points can limit a study's power to detect a significant result. In later analyses, it might be possible to combine data from several studies. When several studies that have small numbers of data points and marginally statistically significant results are combined, statistically significant results can be revealed because the statistical power has been increased (Cohn and Becker 2003; Walker et al. 2008; Haidich 2010). In contrast, Filter 2 could also exclude studies that were well powered but failed to detect an NMDR curve. Statistical issues are also pertinent to Filter 3, which restricts consideration to studies that found a lowest observed-effect level or

lowest observed-adverse-effect level. The identification of such levels can be heavily influenced by study design, particularly when small numbers of animals are tested. That concern is also relevant to Filters 4 (a) and (b). Issues related to study quality are discussed further below.

Evaluating Study Quality

The SOTS evaluation used an expert-based approach to evaluate study quality. Although expert judgment clearly is important, and the committee could find no clear description of a strategy or criteria for assessing the studies used in the evaluation (see Table 2-1). Lack of such criteria and of their systematic application to the studies raises serious concerns about the ability of the SOTS evaluation to reach conclusions regarding the degree to which NMDR curves are evident in the scientific literature. It also compromises the assessment of whether NMDR curves require changes in EPA chemical-testing and risk-assessment strategies. Another National Research Council committee is providing relevant guidance on study-quality issues, such as risk of bias, to EPA's Integrated Risk Information System program (NRC in press). Recommendations from that committee could be supplemented with study-quality criteria specific to evaluating the evidence on NMDR curves.

Evaluation of study quality for the SOTS evaluation should include whether a study has the design elements needed to determine the presence or absence of an NMDR curve. Most standard toxicity-testing protocols have low sensitivity and little statistical power for detecting NMDR curves, particularly at the lower end of the dose–response curve, because they typically test only three or four doses. Thus, it is not surprising that most studies in the SOTS evaluation did not find NMDR curves and that reproducibility of the small number of studies that have shown such an effect is low. Figure 2-1 gives a hypothetical example of the relationship between the statistical power of a toxicology experiment and the ability to reveal NMDR curves.

If few studies have the necessary methodologic features to evaluate whether a dose–response curve shows an NMDR relationship, it will be necessary to consider other studies. Such studies may have methodologic challenges, such as low statistical power, that raise questions about how the data should be analyzed and interpreted. For example, how should issues regarding multiple comparisons (such as inflated type I error rate in all tests combined) be handled in evaluating secondary or tertiary hypotheses? How should detection of a quadratic effect leading to an NMDR curve be handled if the original analysis of variance (ANOVA) was not statistically significant? What statistical methods should be used to perform robust analyses of variances in response that might be nonmonotonic but that might be functions of mean response and have outliers?

Those are difficult and controversial questions, and addressing them goes beyond the present committee's task. In general, the committee supports the conduct of predefined secondary analyses as appropriate in support of evidence

synthesis, such as meta-analysis. Criteria will need to be developed for determining appropriate statistical approaches to evaluate NMDR curves. For example, ANOVA is not designed to detect NMDR curves, so criteria for using such other approaches as biologically based dose–response models, polynomial functions, or splines should be considered.

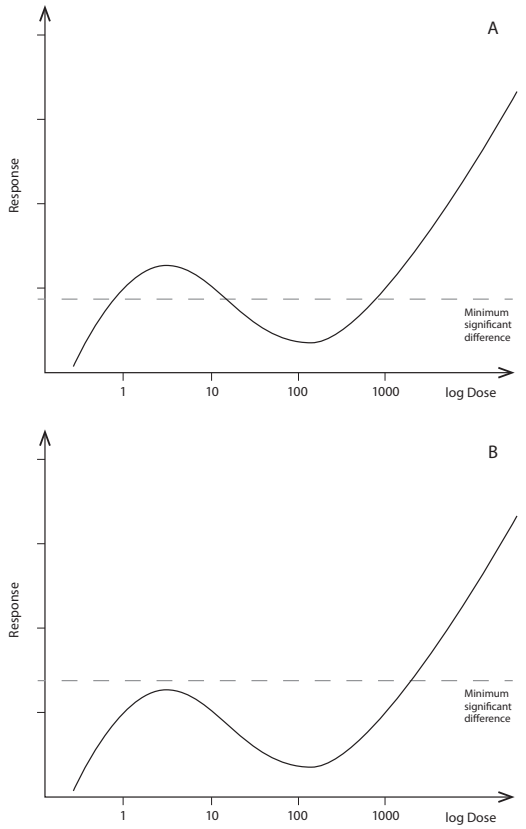

FIGURE 2-1 Relationship between the statistical power of a toxicologic experiment and the ability to reveal a nonmonotonic dose–response (NMDR) relationship. (A) Hypothetical NMDR curve. The horizontal dashed line depicts the minimum significant difference (MSD), the effect magnitude that can be detected as statistically significantly different from untreated controls a high percentage of the time. Effects below this line will be classified as not statistically significant. The magnitude of an MSD depends on several factors, such as the number of animals per dose group and the inherent variability of the measured response. In this example, the experiment is sufficiently powered to reveal an NMDR curve. (B) Same hypothetical NMDR curve as in (A), but with an underpowered experiment, which results in a larger MSD. In this case, an NMDR curve will tend to be overlooked because the nonmonotonicity will be classified as "background fluctuation" with no statistical significance.

Data Presentation and Summarization

Consistency of Exposure Descriptions

Given the wide range of chemicals and doses considered in the SOTS evaluation, environmental exposure ranges for each chemical could be provided to give context to the data. That would help in understanding the similarities and differences in findings between the studies described. Some examples of such context are already in the document; doses tested for genistein in the diet are described as including levels found in human diets (p. 97), and figures are used to demonstrate points of inflection in in vitro and in vivo data compared with environmental exposure concentrations (pp. 52-55). Alternative presentations might also be useful. For example, tabulation of the data might allow easier comparisons of the data. However, the presentation should consider the issues related to statistical power as described above.

It would also be helpful to report exposures to each chemical in the same units throughout a given section (for example, either as micrograms per liter or parts per billion). Inconsistency in the units is particularly problematic in the appendixes, where study tables are reproduced directly from publications. It would be helpful to convert standard international units to conventional units (or vice versa) in such tables. Consideration might also be given to creating a single large table for each appendix (expanded from Tables 4.1-4.4).

The committee recommends that, where it is possible, figures for a given end point be provided to show multiple dose–response relationships on the same scale for ease of comparison (see, for example, NTP 2013).

Consistency of Study Descriptions

Using a consistent format to present data from the individual studies that evaluated NMDR curves would allow multiple authors to contribute while permitting consistency in style and substance; this would make it easier for readers to understand the data and follow the conclusions. Tables 4.1–4.4 in the SOTS evaluation are purported to present the studies that provide evidence of NMDR curves. Some of the variation in how the data were analyzed in the various sections contributes to slightly different formats of the tables. However, the data presented are restricted to study-design elements and provide no information on the specific end points for which EPA asserts that NMDR curves were demonstrated. Such information would make the tables more useful by providing context for the evidence.

The committee found that Appendix A (the section on estrogen disruptors in mammals) provides useful context of the reported NMDR curves. For example, the appendix specifies the total number of end points evaluated in each study, which is then compared with the evidence in other generations in the same study (same doses, sex, and strain). The approach of cross-checking an

NMDR relationship in one generation by looking at the same end point in other generations is logical and valid.

In the thyroid section, the study descriptions of thyroid hormones should distinguish whether measurements are of total or of free triiodothyronine (T3) and thyroxine (T4). Changes in total hormone concentrations may be due simply to alterations in binding and may or may not be associated with alterations in the concentrations of the bioactive free hormones.

Summarization of Data

The SOTS evaluation provides detailed descriptions of the studies considered but little or no synthesis or summarization of the data to compare results among studies or to understand how the evidence was weighed to reach conclusions. The thyroid sections (Sections 4.2.4.4 and 4.2.4.5) do the best job of summarizing the overall findings of the studies that demonstrated NMDR curves. This type of data synthesis and interpretation is lacking in the androgen and estrogen sections. It will be important to ensure that other conclusions embedded throughout the evaluation are adequately supported. For example, support is needed for the statement on p. 75 that "this pattern suggests, perhaps, that NMDRs may be more prevalent in shorter-term assays, especially during periods of system disequilibrium." Another example, on p. 113, inadequately summarizes the evidence on peripubertal exposure to DEHP on male rat reproductive development: "While effects on [preputial separation] and body and reproductive organ weights with an NMDR were observed in the Ge et al. (2007) [B.2.c.3], they were not seen by Noriega et al. (2009) [B.2.c.4] in either of two rat strains studied, and some of the other effects reported in Ge et al. are not consistent with findings from other publications."

Approach to Synthesizing Evidence

The committee evaluated whether the SOTS evaluation has transparently laid out the process for integrating the evidence from various studies and provided a consistent rationale for WOE from various lines of research to support reasonable conclusions. The introduction to the SOTS evaluation describes the WOE used by EPA to determine hazards and risks associated with chemicals. The approach includes "assembling the relevant data; evaluating that data for quality and relevance; and an integration of the different lines of evidence to support conclusions concerning a property of a substance. The significant issues, strengths, and limitations of the data and the uncertainties that deserve serious consideration are presented, and the major points of interpretation highlighted" (p. 27). However, the SOTS evaluation notes that "for sections of [the] review, judgments of likelihood were based on relevant examples rather than a formal [WOE]" (p. 27). Thus, the committee understands that a formal WOE evaluation was not performed but was nonetheless struck by the lack of transparency in how the authors integrated the evidence in each section. Indeed, although such a

synthesis forms the core of what the SOTS evaluation intends to communicate, the balancing and weighing of all the evidence was not described, let alone described in a transparent manner that would readers to reach the same conclusion from the same data.

As advocated in previous National Research Council reports (e.g., NRC 2011), presentation of reviewed studies should be standardized in tabular or graphic form to capture the key dimensions of study characteristics, WOE, and utility for addressing the question under consideration. Transparency and clarity in the lines of evidence considered and how it was integrated to draw conclusions would help to minimize unintended or perceived biases on the part of the authors or the readers of the SOTS evaluation. For example, a table could specify by publication the end points evaluated and the dose–response evidence, which would illustrate studies that did and did not show NMDR curves. Alternatively, separate tables could be created for each end point. Either format would put the mass of the evidence in one place and allow readers to see easily the number of end points and the ones that did and did not have NMDR curves. Using figures that provide the same information for each of the studies that are included in the review would be a consistent and clear way to display the data. Such presentations would allow readers to observe the patterns in the data that led to the authors' conclusions. Examples of approaches that could be used to guide revisions of the SOTS evaluation include those used by NTP's Center for the Evaluation of Risks to Human Reproduction and those being developed by NTP's Office of Health Assessment and Translation (NTP 2013), other proposed data review and synthesis tools recently developed (e.g., Woodruff and Sutton 2011), and approaches recommended by other National Research Council committees (NRC in press).

Statistical methods are available for combining evidence from multiple studies, including meta-analytic approaches (e.g., Greenland and Longnecker 1992; Berlin et al. 1993; Berlin and Coldiz 1999; Steenland et al. 2001; Sutton and Higgins 2008; Orsini et al. 2012) and Bayesian approaches (e.g., Sutton and Abrams 2001). Those methods could be adapted to evaluating the evidence on NMDR curves. However, that would need further research and development before implementation, inasmuch as current methods rely on a common point estimate of a single parameter and its standard error from each study and do not have the capacity to explore more complex relationships. Alternatives to conventional meta-analysis include pooling individual participant data from different studies for modeling or performing Bayesian hierarchic modeling. Regardless of how the methods are adapted, assessment of study heterogeneity, full specification of study design, and accessibility to individual participant data will be important for evidence integration and synthesis of NMDR data.

CONCLUSIONS

The committee focused on whether EPA fairly and soundly evaluated the evidence from diverse sources (in vivo, animal, mode-of-action, and epidemio-

logic studies) and whether the evaluation would be accepted as robust, transparent, objective, and repeatable by the scientific community. EPA has made it clear that time and resource constraints led to its decision to use separate, expert-based evaluations to develop the SOTS evaluation of NMDR curves. However, the agency failed to establish (or enforce) a clear set of methods for collecting and analyzing the evidence on NMDR curves to ensure that the groups conducted their assessments in a clear, consistent, and therefore replicable manner. Instead, the groups determined independently how to perform their analyses. Although such an approach might be appropriate as an internal scoping exercise for the agency, the SOTS evaluation is to be its foundational synthesis of the literature on NMDR curves for the estrogen, androgen, and thyroid pathways and addresses biologic responses that are often counterintuitive. The document will probably be a milestone event in the history of EPA's engagement with endocrine disruptors because it draws conclusions about the existence of NMDR curves and the conditions under which they occur that will to be used to inform decisions about the agency's toxicity-testing strategies and risk-assessment practices. Given its importance and its broad use, the committee judges that the SOTS evaluation should meet a higher standard of evaluation, particularly given the heated controversy surrounding this issue. Methods that provide a more systematic approach and greater transparency are necessary, or it will be too easy to dismiss the analysis as superficial or even biased in the literature selection and evaluation. Although it is clear that the authors spent enormous time and energy in developing the evaluation, it is fundamentally compromised, at least in appearance.

EPA has acknowledged the lack of a consistent and transparent process for data identification, selection, and evaluation and is actively engaged in establishing new procedures on the basis of recommendations from other National Research Council reports (NRC 2009, 2011). An upcoming National Research Council report will address methods specifically for performing evidence evaluations and evidence integration for EPA's Integrated Risk Information System. EPA has already taken steps to address shortcoming in the current SOTS evaluation by developing a Performance Work Statement for subcontractors to conduct systematic literature searches, data extraction, and evaluation of the evidence on NMDR relationships (EPA, unpublished material, September 6, 2013). However, the results of this activity were not available to the committee, so recommendations are restricted to what was presented in the SOTS evaluation.

RECOMMENDATIONS

EPA's SOTS evaluation should be revised to provide more systematic and transparent approaches to evaluating the literature on NMDR curves for the three hormone pathways. Guidance for such approaches is available from clinical sciences (e.g., IOM 2011), other National Research Council reports (e.g., NRC 2011), and those being developed at other government agencies (e.g., NTP 2013). Important considerations include the following:

- The mismatch between the breadth of the three central scientific questions of the SOTS evaluation and the narrower scope of the analytic approach used to address them should be rectified by either narrowing the scope of the questions or broadening the scope of the analysis. Specific important issues include these:

    - The decision to exclude human studies should be reconsidered, particularly because the analysis will ultimately be used to make decisions about human health risk-assessment practices.
    - The MOAs considered for each of the hormonal pathways should be clarified and considered in the context of the breadth of the questions to be answered.

- An analytic plan should be developed and applied consistently to the evidence on the three hormone pathways. Important elements of the plan include predefining and documenting the literature-search strategies and their results, criteria for selecting studies for analysis, criteria for determining study quality, templates for presenting evidence consistently in tabular and graphic form, and approaches to integration of evidence. The following are specific consideration for these elements:

    - Improve the justification and context for the definitions of *low-dose effect, resilience,* and *adverse effect* used in the SOTS evaluation.
    - The methodologic features that would be necessary for a study to be able to detect an NMDR relationship should be established. Ideally, studies would have multiple dose groups that were spaced across a defined exposure domain, including doses below those typically tested. Statistical design, biologic plausibility, and replicability should be factored into interpreting and weighing the evidence from such studies.
    - Study exclusion and inclusion criteria should be established. Although statistical significance is an important consideration, it should not be an absolute criterion for including or excluding studies, inasmuch as standard toxicity-testing strategies generally do not have sufficient sensitivity and statistical power to detect NMDR curves.
    - Study quality criteria should be established. Statistical criteria should be given particular attention. It will be important to balance study-quality criteria that are based on statistical significance and those based on biologic plausibility.
    - Secondary analyses of other studies may be necessary. Current methods for performing post hoc analysis of data and for combining evidence from multiple studies might be adapted for such purposes but will require research and development before implementation. EPA should consider soliciting input from the biostatistics community on the best methods to pursue in the long term and on what measures should be taken to complete the SOTS evaluation.

- Evidence tables and graphic presentations should use consistent units for varied studies (when possible), present multiple dose–response curves on the same scale (when possible) to facilitate comparisons, and provide more context for exposure ranges.
- The evidence on each hormone pathway should be summarized and synthesized to document the key evidence and WOE analysis that led to conclusions.

REFERENCES

Abrami, P.C., P.A. Cohen, and S. d'Apollonia. 1988. Implementation problems in meta-analysis. Rev. Educ. Res. 58(2):151-179.

Andersen, M.E., J.E. Dennison, R.S. Thomas, and R.B. Conolly. 2005. New directions in incidence-dose modeling. Trends Biotechnol. 23(3):122-127.

Berlin, J.A., and G.A. Colditz. 1999. The role of meta-analysis in the regulatory process for foods, drugs, and devices.JAMA 281(9):830-834.

Berlin, J.A., M.P. Longnecker, and S. Greenland. 1993. Meta-analysis of epidemiologic dose-response data. Epidemiology 4(3):218-228.

Birnbaum, L.S., K.A. Thayer, J.R. Bucher, and M.S. Wolfe. 2013. Implementing systematic review at the National Toxicology Program: Status and next steps. Environ. Health Perspect. 121(4):A108-A109.

Cohn, L.D., and B.J. Becker. 2003. How meta-analysis increases statistical power. Psychol. Methods 8(3):243-253.

Dolinoy, D.C., D. Huang, and R.L. Jirtle. 2007. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc. Natl. Acad. Sci. USA 104(32):13056-13061.

EFSA (European Food Safety Authority). 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):3121.

EPA (U.S. Environmental Protection Agency). 2013a. 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: http://epa.gov/ncct/download_files/edr/NMDR.pdf [accessed Aug. 12, 2013].

EPA (U.S. Environmental Protection Agency). 2013b. Part 1. Status of Implementation of Recommendations. Materials Submitted to the National Research Council, by Integrated Risk Information System Program, U.S. Environmental Protection Agency, January 30, 2013[online]. Available: http://www.epa.gov/IRIS/iris-nrc.htm [accessed Oct. 22, 2013].

EPA (U.S. Environmental Protection Agency). 2014. Integrated Risk Information (IRIS) Glossary [online]. Available: http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do?details=&glossaryName=IRIS%20Glossary [accessed Feb. 19, 2014].

Filardo, E.J., and P. Thomas. 2012. Minireview: G protein-coupled estrogen receptor-1, GPER-1: Its mechanism of action and role in female reproductive cancer, renal and vascular physiology. Endocrinology 153(7):2953-2962.

Ge, R.S., G.R. Chen, Q. Dong, B. Akingbemi, C.M. Sottas, M. Santos, S.C. Sealfon, D.J. Bernard, and M.P. Hardy. 2007. Biphasic effects of postnatal exposure to diethylhexylphthalate on the timing of puberty in male rats. J. Androl. 28:513-520 (as cited in EPA 2013a).

Greenland, S., and M.P. Longnecker. 1992. Methods for trend estimation from summarized dose-response data, with application to meta-analysis. Am. J. Epidemiol. 135(11):1301-1309.

Guerrero-Bosagna, C., T.R. Covert, M.M. Haue, M. Settles, E.E. Nilsson, M.D. Anway, and M.K. Skiinner. 2012. Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers. Reprod. Toxicol. 34(4):694-707.

Guyatt, G., A.D. Oxman, E.A. Akl, R. Kunz, G. Vist, J. Brozek, S. Norris, Y. Falck-Ytter, P. Glasziou, H. DeBeer, R. Jaeschke, D. Rind, J. Meerpohl, P. Dahm, and H.J. Schünemann. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J. Clin. Epidemiol. 64(4):383-394.

Haidich, A.B. 2010. Meta-analysis in medical research. Hippokratia 14(suppl. 1):29-37.

Higgins, J.P.T., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, The Cochrane Collaboration [online]. Available: http://handbook.cochrane.org/ [accessed Feb. 7, 2014].

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

Keller, D.A., D.R. Juberg, N. Catlin, W.H. Farland, F.G. Hess, D.C. Wolf, and N.G. Doerrer. 2012. Identification and characterization of adverse effects in the 21st century toxicology. Toxicol. Sci. 126(2):291-297.

Koustas, E., J. Lam, P. Sutton, P.I. Johnson, D. Atchley, S. Sen, K.A. Robinson, D. Axelrad, and T.J. Woodruff. In press. Applying the navigation guide: Case study #1. The impact of developmental exposure to perfluorooctanoic acid (PFOA) on fetal growth. A systematic review of the non-human evidence. Environ. Health Perspect.

Kundakovic, M., K. Gudsnuk, B. Franks, J. Madrid, R.L. Miller, F.P. Perera, and F.A. Champagne. 2013. Sex-specific epigenetic disruption and behavioral changes following low-dose in utero bisphenol A exposure. Proc. Natl. Acad. Sci. USA. 10(24):9956-9961.

Kushman, M.E., A.D. Kraft, K.Z. Guyton, W.A. Chiu, S.L. Makris, and I. Rusyn. 2013. A systematic approach for identifying and presenting mechanistic evidence in human health assessments. Regul. Toxicol. Pharmacol. 67(2):266-277.

Manikkam, M., R. Tracey, C. Guerrero-Bosagna, and M.K. Skinner. 2013. Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. PLoS One 8(1):e55387.

Meline, T. 2006. Selecting studies for systematic review: Inclusion and exclusion criteria. CICSD 33(Spring):21-27.

Noriega, N.C., K.L. Howdeshell, J. Furr, C.R. Lambright, V.S. Wilson, and L.E. Gray, Jr. 2009. Pubertal administration of DEHP delays puberty, suppresses testosterone production, and inhibits reproductive tract development in male Sprague-Dawley and Long-Evans rats. Toxicol. Sci. 111(1):163-178 (as cited in EPA 2013a).

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

NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

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

NRC (National Research Council). 2013. Critical Aspects of EPA's IRIS Assessment of Inorganic Arsenic: Interim Report. Washington, DC: The National Academies Press.

NRC (National Research Council). In press. Review of Revisions to the 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 Program, U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC [online]. Available: http://www.epa.gov/endo/pubs/edmvs/lowdosepeerfinalrpt.pdf (as cited in EPA 2013a).

NTP (National Toxicology Program). 2013. OHAT Implementation of Systematic Review [online]. Available: http://ntp.niehs.nih.gov/go/38673 [accessed Oct. 23, 2013].

Orsini, N., R. Li, A. Wolk, P. Khudyakov, and D. Spiegelman. 2012. Meta-analysis for linear and nonlinear dose response relations: Examples, an evolution of approximations, and software. Am. J. Epidemiol. 175(1):66-73.

Prins, G.S., W.Y. Tang, J. Belmone, and S.M. Ho. 2008. Developmental exposure to bisphenol A increases prostate cancer susceptibility in adult rats: Epigenetic mode of action is implicated. Fertil. Steril. 89(2 Suppl):e41. doi: 10.1016/j.fertnstert.2007.12.023.

Somm, E., C. Stouder, and A. Paoloni-Giacobino. 2013. Effect of developmental dioxin exposure on methylation and expression of specific imprinted genes in mice. Reprod. Toxicol. 35:150-155.

Steenland, K., A. Mannetje, P. Bofetta, L. Stayner, M. Attfield, J. Chen, M. Dosemeci, N. DeKlerk, E. Hnizdo, R. Koskela, and H. Checkoway. 2001. Pooled exposure-response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: An IARC multicentre study. Can. Caus. Control. 12(9):773-784.

Sutton, A.J., and K.R. Abrams. 2001. Bayesian methods in meta-analysis and evidence synthesis. Stat. Methods Med. Res. 10(4):277-303.

Sutton, A.J., and J.P.T. Higgins. 2008. Recent developments in meta-analysis. Stat. Med. 27:625-650.

Tang, W.Y., L.M. Morey, Y.Y. Cheung, L. Birch, G.S. Prins, and S.M. Ho. 2012. Neonatal exposure to estradiol/bisphenol A alters promoter methylation and expression of Nsbp1 and Hpcal1 genes and transcriptional programs of Dnmt3a/b and Mbd2/4 in the rat prostate gland throughout life. Endocrinology. 53(1):42-55.

Teeguarden, J.G., and S. Hason-Drury. 2013. A systematic review of bisphenol A "low dose" studies in the context of human exposure: A case for establishing standards for reporting "low-dose" effects of chemicals. Food Chem. Toxicol. 62:935-948.

Vandenberg, L.N., T. Colborn, T.B. Hayes, J.J. Heindel, D.R. Jacobs, D.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.

Walker, E., A.V. Hernandez, and M.W. Kattan. 2008. Meta-analysis: Its strengths and limitations. Cleve. Clin. J. Med. 75(6):431-439.

Wolstenholme, J.T., E.F. Rissman, and J.J. Connelly. 2011. The role of bisphenol A in shaping the brain, epigenome and behavior. Horm. Behav. 59(3):296-305.

Woodruff, T.J., and P. Sutton. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff. 30(5):931-937.

Woodruff, T. J., and P. Sutton. In press. The navigation guide: An improved method for translating environmental health science into better health outcomes. Environ. Health Perspect.

Woodruff, T.J., L. Zeise, D.A. Axelrad, K.Z. Guyton, S. Janssen, M. Miller, G.G. Miller, J.M. Schwartz, G. Alexeeff, H. Anderson, L. Birnbaum, F. Bois, V.J. Cogliano, K. Crofton, S.Y. Euling, P.M. Foster, D.R. Germolec, E. Gray, D.B. Hattis, A.D. Kyle, R.W. Luebke, M.I. Luster, C. Portier, D.C. Rice, G. Solomon, J. Vandenberg, and R.T. Zoeller. 2008. Meeting report: Moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. Environ. Health Perspect. 116(11):1568-1575.

3

The Environmental Protection Agency's Evaluation of Implications of Nonmonotonic Dose–Response Relationships for Current Toxicity-Testing Strategies and Risk-Assessment Practices

This chapter reviews the approaches used by the US Environmental Protection Agency (EPA) to answer questions about the adequacy of its toxicity-testing strategies and risk-assessment practices on the basis of its findings about nonmonotonic dose response (NMDR) relationships for endocrine disruptors in its draft *State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures* (SOTS evaluation). The questions that EPA sought to answer are presented in Box 3-1. The committee evaluated whether EPA applied sound scientific principles, whether potential effects on toxicity-testing guidelines and risk-assessment practices were appropriately considered, and whether conclusions were adequately justified in response to Questions 2 and 3.

The answers to those questions depend on the scope of and answer to Question 1. As discussed at length in Chapter 2, the SOTS evaluation did not provide adequate analysis and support for its conclusions that NMDR curves were found more often in in vitro studies, at high doses, and for exposures of short duration. The committee found that EPA's approach to evaluating whether NMDR curves exist for endocrine disruptors was not systematic, consistent, or transparent, especially with regard to the literature on estrogen and androgen effects. Thus, the same issues and problems discussed in Chapter 2 carry over to the review of the evidence in this chapter.

THE ENVIRONMENTAL PROTECTION AGENCY'S EVALUATION OF THE ADEQUACY OF TOXICITY TESTING

In trying to answer Question 2, EPA posed three more specific questions:

2a. Are there adverse effects with NMDRs that are not being identified by using the current chemical testing strategies?

2b. Are there NMDRs for adverse effects below the no observed adverse effect levels (NOAELS) or benchmark doses (BMD) derived from the current testing strategies?

2c. Do EPA chemical testing strategies detect relevant adverse effects for chemicals which produce NMDR curves for specific endpoints?

No description of "current chemical testing strategies" is provided, so it was unclear to the committee whether they included EPA's guidelines for toxicity testing under the Toxic Substances Control Act (TSCA) or the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), testing associated with the Endocrine Disruptor Screening Program (EDSP), testing associated with EPA's ToxCast program, or the larger body of investigative toxicology and clinical or epidemiologic research. Each of those testing strategies has specific goals and applications that are important for identifying potential hazards, including those associated with endocrine disruptors, and for framing the mode of action (MOA) or weight of evidence approaches that ultimately affect risk assessments. For the purposes of the present review, the committee assumed that EPA's questions and conclusions pertain to TSCA, FIFRA, and EDSP testing strategies.

EPA cited the work of a few European agencies, such as the European Food Safety Authority (EFSA 2013) and the Danish Centre on Endocrine Disrupters (Hass et al. 2013), that have analyzed the strengths and weaknesses of testing guidelines for addressing endocrine-disruptor toxicity and dose–response issues. Their approaches are relevant to the SOTS evaluation because they explicitly analyze the adequacy of specific testing methods. EPA should use a similar approach to analyze the strengths and weaknesses of its testing guidelines for addressing its questions about their adequacy for detecting adverse effects associated with NMDR curves.

BOX 3-1 Central Scientific Questions EPA Addressed in the SOTS Evaluation (EPA 2013)

- **Question 1**:[a] Do [NMDRs] exist for chemicals and if so under what conditions do they occur?
- **Question 2**: Do NMDRs capture adverse effects that are not captured using [EPA's] current chemical testing strategies (i.e., false negatives)?
- **Question 3**: Do NMDRs provide key information that would alter EPA's current weight of evidence conclusions and risk assessment determinations, either qualitatively or quantitatively?

[a]See Chapter 2 for committee's evaluation of how Question 1 was answered.

Failure to present the testing strategies under consideration and to develop a framework for analyzing them in the context of the NMDR evidence makes it difficult to understand how EPA drew its conclusions. As noted in Chapter 2, the SOTS evaluation should meet a higher standard of evaluation if it is to be a foundational synthesis of the literature to inform decisions about EPA's testing strategies. In the sections below, the committee considers issues regarding each of the individual questions that EPA sought to answer.

Question 2a: Are There Adverse Effects with NMDRs That Are Not Being Identified by Using the Current Chemical Testing Strategies?

In response to this question, the SOTS evaluation states that EPA's toxicity-testing strategy is not intended to identify all possible adverse outcomes of chemical exposure but instead is intended to provide broad coverage of sensitive end points that are relevant to human health. It states that estrogen-, androgen-, and thyroid-responsive effects are evaluated as part of testing in EDSP and that multigeneration studies include end points that are sensitive to androgen and estrogen action. It concludes that "current testing strategies are unlikely to mischaracterize, as a consequence of NMDR, a chemical that has the potential for adverse perturbations of the estrogen, androgen or thyroid pathways." However, the SOTS evaluation provides no analysis to support that conclusion.

Answering Question 2a in a defensible and transparent manner requires a comparison of the end points evaluated in current testing strategies that are thought to be sensitive to estrogen, androgen, or thyroid effects with the end points associated with NMDR curves. The MOAs associated with those effects should then be evaluated in the context of additional supportive data, if available, to help in understanding the sequences of key events that ultimately lead to the adverse effects. As the SOTS evaluation noted, not all measured end points are themselves adverse or necessarily key events associated with an adverse response, so careful consideration is needed. Early development is an especially sensitive window of susceptibility, and perturbations that are adaptive in the average adult may give rise to an adverse response in the fetus, for example. The analysis should give careful consideration to the suite of end points assessed in reproductive and developmental toxicity and related tests, given the importance of hormones in fetal, perinatal, and adolescent development. A more rigorous approach to analyzing the evidence would help in framing the data to determine whether current testing strategies are qualitatively identifying hazards irrespective of whether an NMDR curve is present and whether follow-on investigative studies would be needed to differentiate key events associated with adverse events from events that are considered adaptive.

Question 2b. Are there NMDRs for Adverse Effects Below the No Observed Adverse Effect Levels or Benchmark Doses Derived From the Current Testing Strategies?

In response to this question, the SOTS evaluation states that "for estrogen, androgen or thyroid MoA[s] that provide adequate information to make an assessment, our evaluation shows that there is not sufficient evidence of NMDRs for adverse effects below the NOAELs or BMD derived from the current testing strategies." However, the evaluation provides no analysis of EPA's testing protocols for their adequacy to detect NMDR curves for adverse responses at doses below NOAELs or BMDs derived from traditional toxicology-testing protocols. Without such an analysis of the strengths and weaknesses associated with testing protocols, EPA's conclusions are not well supported.

Answering Question 2b in a more rigorous manner will require an analysis of protocols for their ability to detect adverse effects associated with perturbations of the estrogen, androgen, and thyroid pathways and for their ability to detect NMDR curves. EPA should pay particular attention to issues of study design—such as the number of dose groups, dose spacing, numbers of animals tested, end points evaluated, and statistical power—in considering whether current testing strategies have the ability to detect NMDR curves for adverse responses. For example, toxicity studies that test the minimum of three dose groups and a control group are inadequate for identifying complex dose–response relationships. That limitation is acknowledged in the SOTS evaluation, as evidenced by EPA's focus on studies that use four or preferably more dose groups. As discussed in Chapter 2, EPA was left with a relatively small number of studies from which to draw definitive conclusions on NMDR curves for adverse effects associated with a broad range of exposures and multiple potential outcomes. That may well reflect true inadequacies in the literature and even in EPA's testing protocols, but a more rigorous analysis of the literature is needed to support any conclusions, including a finding that the available data are inadequate to reach a conclusion.

Question 2c. Do EPA Chemical Testing Strategies Detect Relevant Adverse Effects for Chemicals Which Produce NMDR for Specific Endpoints?

EPA's response to Question 2c appears to be that current testing assays in the EDSP's screening battery detect disruption of estrogen, androgen, and thyroid pathways and that standard multigenerational tests "have measures that are sensitive to disruption of the estrogen and androgen signaling pathways" (p. 13). However, as discussed in relation to Questions 2a and 2b, the SOTS evaluation provides no analysis to support these answers. A rigorous comparison of the end points evaluated in EPA's current testing strategy with the expected outcomes of exposures to chemicals that affect the estrogen, androgen, and thyroid pathways is needed. EPA could then identify the outcomes for which NMDR curves are

plausible and consider the extent to which they would be captured by the other end points. The analysis could then be supported by a review of the human physiologic, toxicologic, and epidemiologic literature to provide context for toxicity assessments conducted in laboratory animals.

It is unclear whether "relevant adverse effects" refers to ones that would be important for the purposes of hazard identification or ones that would lead to further testing requirements or investigative research to establish adversity. Furthermore, not all end points are considered equal in a risk assessment; some effects are given more weight than others. For example, cardiac malformations can provide stronger evidence of a developmental hazard than would modulation of a hormone concentration. Both are measurable end points, but they have different considerations in hazard identification. The SOTS evaluation does not include an analysis that considered both the extent of end-point coverage and the quality of the response.

THE ENVIRONMENTAL PROTECTION AGENCY'S EVALUATION OF THE IMPLICATIONS OF NONMONOTONIC DOSE–RESPONSE RELATIONSHIPS FOR RISK-ASSESSMENT PRACTICES

Question 3 of the SOTS evaluation was: Do NMDRs provide key information that would alter EPA's current weight of evidence conclusions and risk assessment determinations, either qualitatively or quantitatively? A conclusion was drawn that "data from studies in which NMDRs are identified may be biologically relevant and as such should be evaluated in context with the totality of the available scientific data in weight of evidence (WoE) conclusions and risk assessment determinations. . . . NMDRs can have impact on both qualitative and quantitative risk assessments, but cannot be considered in isolation from other data for the chemical and biological response being considered" (p. 14).

Although risk-assessment guidelines specific to endocrine disruptors have not been developed, the committee recognizes that endocrine disruption has been treated as a specific MOA in guideline categories, such as those for developmental toxicity (EPA 1991), reproductive toxicity (EPA 1996), cancer (EPA 2005a,b), neurotoxicity (EPA 1998a), mutagenicity (EPA 1986), and ecologic effects (EPA 1998b). EPA acknowledged that many prior reviews and expert panels concluded that there is a need to address specifically how endocrine disruptor–related dose–response relationships affect current risk-assessment practices. However, EPA did not analyze how evidence of NMDR curves would be considered in its current risk-assessment practices and drew only the general conclusion that NMDR curves, if deemed important and relevant, would have an effect. Thus, the SOTS evaluation fails to provide the analysis needed to answer Question 3 in a defensible manner.

To address Question 3 directly and respond to the challenges that have been raised, the committee recommends that the draft SOTS evaluation be revised to include a specific analysis and presentation of how NMDR relation-

ships, to the extent that evidence supports their existence for adverse effects, would be addressed in current risk-assessment guidelines and practices and under what circumstances they would be found inadequate. A key consideration will be an examination of how NMDR curves bear on the WOE and MOA analyses that drive the risk-assessment process. Even if the current literature is found not to provide compelling evidence of NMDR curves of health relevance in the exposure range of regulatory concern, an assessment of how such evidence would alter risk assessment is needed to answer Question 3. For example, consideration could be given to whether cross-species extrapolations should be performed differently from current practices. Those types of analyses are needed to determine whether NMDR curves provide key information that would alter EPA's current WOE conclusions and risk-assessment determinations.

Noticeable omissions in the draft SOTS evaluation are consideration of human studies—including the physiologic, clinical, and epidemiologic evidence—and discussion of how evidence of NMDR curves in ranges relevant to humans would be incorporated into risk assessments. Furthermore, little attention is paid to potential issues associated with, for example, cross-species, lifestage, sex, route-to-route, and in vitro–to–in vivo extrapolations that help to frame the MOA and dose–response analyses that are important for developing human health risk assessments. Those issues can be even more problematic in extrapolating from aquatic or ecologic studies to the broader ecosystem or human health. Clarifying how such issues could be addressed in human and environmental health assessments would be an important contribution to the SOTS evaluation.

Thus, the committee recommends that the answer to Question 3 be justified with specific reference to EPA’s risk-assessment guidelines. As noted earlier, EPA could use examples from other organizations to guide its own analyses. For example, the Danish Centre on Endocrine Disrupters (Hass et al. 2013) provides specific examples of the strengths and weaknesses of European testing guidelines, establishment of dose–response relationships (BMDs) or NOAELs, and statistical considerations and uncertainties for use in risk assessment. EPA should consider and expand on such analyses to establish the bridge between the state of the science for NMDR curves associated with estrogen, androgen, and thyroid pathways and how they could be implemented in risk assessments in a transparent manner that is ultimately useful to the regulatory community. In that way, the final SOTS evaluation would provide a more coherent and complete response to the questions that are posed.

As EPA considers this recommendation, it should also consider reiterating and clarifying its operating definitions. In addition to the definitions of low-dose effect, resilience, and adverse effect discussed in Chapter 2, definitions of WOE and MOA in a risk-assessment context should be clarified. Clear definitions are especially important so that future debates can be based on common understandings of the strengths and weaknesses of the science that underpins methods and results.

CONCLUSIONS

The committee found that the SOTS evaluation provided little documentation of the analyses performed to answer questions about toxicity-testing strategies and risk-assessment practices. EPA stated in the Executive Summary of the SOTS evaluation that it did not intend to review specific testing methods or risk assessments, but this decision needs to be reconciled with the central scientific questions that are posed. Because the SOTS evaluation is to be a foundational synthesis of the literature on NMDR curves for the estrogen, androgen, and thyroid pathways, a higher standard of evaluation is required. Developing at least a simple framework that identifies the testing methods and evaluates their ability to detect endocrine-mediated adverse effects and NMDR relationships is necessary to answer, fully and transparently, questions about the ability of current testing strategies to characterize hazards related to endocrine disruptors appropriately. EPA's conclusions about the adequacy of toxicity testing should be rooted in specific reference to and demonstrations with its own testing guidelines in a systematic manner for the estrogen, androgen, and thyroid pathways even if it eventually decides that the available data are inadequate to reach a conclusion. Similarly, it should devise a framework for evaluating the qualitative and quantitative implications of NMDR curves for WOE conclusions and risk-assessment determinations. In that way, all the factors (including issues, strengths, and weaknesses) associated with establishing dose–response relationships, characterizing MOAs, and understanding route-to-route, in vitro to in vivo, and cross-species extrapolations can be presented openly and transparently, assumptions can be clearly stated, and uncertainties can be acknowledged.

RECOMMENDATIONS

The scope of Questions 2 and 3 should be better defined and tailored to the scope of Question 1 (see Chapter 2). Important elements include the following:

- EPA should specify the toxicity-testing strategies under consideration and evaluate them for their ability to detect NMDR relationships mediated by the estrogen, androgen, and thyroid hormone pathways. The tests should be evaluated in terms of the design criteria that would be necessary to detect such relationships and the adverse effects of interest. The evaluation should include careful consideration of reproductive and developmental toxicity and related tests, given the importance of hormones in fetal, perinatal, and adolescent development.
- As noted in Chapter 2, resilience, or adaptation, is a controversial topic and is not adequately addressed by EPA in the SOTS evaluation. EPA should distinguish between adverse effects and adaptive effects. Because effects that are adaptive in some people are adverse in others, consideration should be given to potential windows of susceptibility (for example, during fetal development),

particularly sensitive populations (for example, those with pre-existing health conditions), and other factors (such as multiple chemical exposures) in making these determinations.

- EPA's SOTS evaluation should be expanded to indicate how NMDR relationships for estrogen, androgen, and thyroid pathways would be addressed under EPA's current risk-assessment guidelines and practices, including statistical considerations, uncertainty analyses, lifestage or susceptibility issues, and MOAs.
- As recommended in Chapter 2, EPA should reconsider its decision to exclude human evidence. Human physiologic, clinical, and epidemiologic literature may be useful in identifying end points, dose–response relationships, and MOAs that are relevant to the estrogen, androgen, and thyroid pathways. If such data are included, they could be used to highlight species differences in chemical interactions with respect to the three hormone pathways, including pharmacokinetic and pharmacodynamics processes that are important considerations in conducting risk assessments.

REFERENCES

EFSA (European Food Safety Authority). 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):3121.

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for Mutagenicity Risk Assessment. EPA/630/R-98/003. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. September 1986 [online]. Available: http://www.epa.gov/raf/publications/pdfs/MUTAGEN2.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 1991. Guidelines for Developmental Toxicity Risk Assessment. EPA/600/FR-91/001. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. December 1991 [online]. Available: http://www.epa.gov/raf/publications/pdfs/DEVTOX.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 1996. Guidelines for Reproductive Toxicity Risk Assessment. EPA/630/R-96/009. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. October 1996 [online]. Available: http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 1998a. Guidelines for Neurotoxicity Risk Assessment. EPA/630/R-95/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. April 1998 [online]. Available: http://www.epa.gov/raf/publications/pdfs/NEUROTOX.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 1998b. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. April 1998 [online]. Available: http://www.epa.gov/raf/publications/pdfs/ECOTXTBX.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 2005a. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. March 2005 [online]. Available: http://www.epa.

gov/raf/publications/pdfs/CANCER_GUIDELINES_FINAL_3-25-05.PDF [accessed Sept. 19, 2013].

EPA (U.S. Environmental Protection Agency). 2005b. Supplemental Guidance for Assessing Susceptibility for Early-Life Exposures to Carcinogens. EPA/630/R-03/003F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. March 2005 [online]. Available: http://www.epa.gov/raf/publications/pdfs/childrens_supplement_final.pdf [accessed Sept. 19, 2013].

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: http://epa.gov/ncct/download_files/edr/NMDR.pdf [accessed Aug. 12, 2013].

Hass, U., S. Christiansen, M. Axelstad, K. Dreisig Sorensen, and J. Boberg. 2013. Input for the REACH-review in 2013 on Endocrine Disrupters. Final report. Center on Endocrine Disrupters, Division of Toxicology and Risk Assessment, National Food Institute, Technical University of Denmark. March 21, 2013 [online]. Available: http://www.mst.dk/NR/rdonlyres/B865F94A-54C0-43DF-AB0F-F07A593E1FD0/0/ReachreviewrapportFINAL21March.pdf [accessed July 15, 2013].

Appendix

Biosketches of the Committee to Review EPA's Draft State of the Science Paper on Nonmonotonic Dose Response

David A. Savitz *(Chair)* is professor of epidemiology and obstetrics and gynecology at Brown University. His primary research interests are in reproductive and environmental epidemiology. He has conducted studies of the causes of adverse pregnancy outcomes, including miscarriage and preterm birth, and pregnancy complications, birth defects, and child health problems. His environmental interests include a wide array of chemical and physical exposures, including exposures to perfluorinated chemicals, pesticides, drinking-water treatment byproducts, and nonionizing radiation. Dr. Savitz was president of the Society for Epidemiologic Research and the Society for Pediatric and Perinatal Epidemiologic Research. He is on the Executive Council of the International Society for Environmental Epidemiology. He has served on several Institute of Medicine (IOM) and National Research Council committees, most recently on the IOM Committee on Obesity Prevention Policies for Young Children. Dr. Savitz was elected to IOM in 2007. He received his PhD in epidemiology from the University of Pittsburgh Graduate School of Public Health.

Andrea Baccarelli is the Mark and Catherine Winkler Associate Professor in the Department of Environmental Health of the Harvard School of Public Health. His Laboratory of Human Environmental Epigenetics investigates molecular mechanisms and biomarkers that reflect reprograming of health and disease trajectories in response to environmental exposures. He is particularly interested in the epigenome and mitochondriome at different life stages. Recent and current projects investigate effects of exposure to particulate air pollution, metals, bisphenol A, phthalates, and pesticides. Dr. Baccarelli received his MD from the University of Perugia, his MPH from the University of Turin, and his PhD from the University of Milan in Italy.

Robert E. Chapin is a senior research fellow at Pfizer, Inc., working in the Developmental and Reproductive Toxicology Center of Expertise. The center helps to interpret and weigh reproductive and developmental data for the organization, is involved in the development and application of in vitro models to screen

compounds, and conducts mechanistic work in developmental and reproductive toxicology. He was the recipient of the Pfizer Achievement Award in 2010. Before joining Pfizer in 2002, Dr. Chapin ran the reproductive toxicology laboratory for the National Toxicology Program at the National Institute of Environmental Health Sciences. He received the National Institutes of Health Director's Award for Scientific Merit in 1995. Dr. Chapin received his PhD in pharmacology from the University of North Carolina at Chapel Hill.

Richard A. Corley is laboratory fellow in the systems toxicology group at the Pacific Northwest National Laboratory operated by Battelle for the US Department of Energy. He specializes in the development of physiologically based pharmacokinetic models and multiscale computational fluid-dynamics–based models of the respiratory system. He has published numerous peer-reviewed papers in toxicology, pharmacokinetic modeling, and cross-species dosimetry applications in human health risk assessments. Dr. Corley was member of the National Research Council Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, Committee on Risk Analysis Issues and Reviews, and Committee to Assess the Health Implications of Perchlorate Ingestion. He received his PhD in environmental toxicology and veterinary biosciences from the University of Illinois at Urbana-Champaign.

George P. Daston is Victor Mills Society Research Fellow at the Procter & Gamble Company and an adjunct professor of pediatrics at the University of Cincinnati. His current research efforts are in toxicogenomics and mechanistic toxicology, particularly in addressing how findings in these fields can improve risk assessment of chemicals and the development of nonanimal alternatives. Dr. Daston has served as president of the Teratology Society, councilor of the Society of Toxicology, member of the Environmental Protection Agency's Science Advisory Board, member of the National Toxicology Program's Board of Scientific Counselors, and member of the National Children's Study Advisory Committee. He has also served on several National Research Council committees, most recently as a member of the Committee on Use of Emerging Science for Environmental Health Decisions. Dr. Daston has been awarded the Distinguished Service Award by the Teratology Society and the George H. Scott Award by the Toxicology Forum and is an elected fellow of the American Association for the Advancement of Science. He received his PhD from the University of Miami.

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 of the Harvard 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 National Research Council and Institute of Medicine committees, including the Committee on the Health Risks of Phthalates. He was chair of the Environment and Reproduction Special Interest Group of the American Society for Reproductive Medicine and serves on the US Consumer Product Safety Commission's Chronic Hazard Advisory Panel that is examining the effects of phthalates on children's health. He received his MD from Albert Einstein College of Medicine and his MPH and ScD from the Harvard School of Public Health. He is board-certified in occupational medicine.

Amy H. Herring is professor of biostatistics and associate chair of the Department of Biostatistics of the University of North Carolina at Chapel Hill. Her research interests include longitudinal and multivariate data, hierarchic models, latent variables, Bayesian methods, reproductive and environmental epidemiology, and maternal and child health. She is a former president of the Eastern North American Region Committee of the International Biometric Society. Dr. Herring is a fellow of the American Statistical Association and serves on the Board of Directors of the International Society for Bayesian Analysis. She received the Mortimer Spiegelman Award for outstanding public-health statistician under the age of 40 years from the American Public Health Association in 2012. Dr. Herring received her ScD in biostatistics from Harvard University.

Andreas Kortenkamp is professor of human toxicology at the Institute for the Environment of Brunel University, London (United Kingdom). His research interests are in exploring environmental pollutants and their combined effects on endocrine diseases. In numerous publications, he has investigated the effects of mixtures of chemicals that can disrupt hormone action. Dr. Kortenkamp has been involved in the development of several reports for the European Commission, including the State of the Art Report on Mixture Toxicology and the State of the Art Assessment of Endocrine Disruptors. He was a member of the World Health Organization–United Nations Environmental Programme panel for evaluating the state of the science of endocrine disruption in 2012. Dr. Kortenkamp was a member of the National Research Council Committee on the Health Risks of Phthalates and is a member of the US Consumer Product Safety Commission's Chronic Hazard Advisory Panel that is examining the effects of phthalates on children's health. He earned his PhD from Bremen University in Germany.

Heather B. Patisaul is an associate professor in the Department of Biology of the North Carolina State University. Her research examines the steroid-dependent mechanisms through which sexually dimorphic behaviors and brain circuits arise, and she also explores the mechanisms by which sexually dimorphic systems and behaviors can be disrupted by environmental estrogens. Her laboratory is interested in the mechanisms by which exposure to environmental estrogens can advance puberty and impair fertility in females. Dr. Patisaul served on the World Health Organization expert panel that assessed the risks of bisphenol A in 2010. She received her PhD from Emory University.

Elizabeth N. Pearce is associate professor of medicine at the Boston University School of Medicine in the Section of Endocrinology, Diabetes, and Nutrition. Her research interests include the sufficiency of dietary iodine in the United States, thyroid function in pregnancy and lactation, thyroid effects of exposure to environmental perchlorate and other potential endocrine disruptors, and the cardiovascular effects of subclinical thyroid dysfunction. She has served as a member of the Board of Directors of the American Thyroid Association (ATA) and is on the Management Council of the International Council for the Control of Iodine Deficiency Disorders Global Network. Dr. Pearce is the recipient of ATA's 2011 Van Meter Lecture Award. She received her MD from Harvard Medical School and her MSc in epidemiology from the Boston University School of Public Health.

Tracey J. Woodruff is professor and director of the Program on Reproductive Health and the Environment in the Department of Obstetrics, Gynecology, and Reproductive Sciences of the University of California, San Francisco. Before joining the university, she was a senior scientist at the US Environmental Protection Agency's National Center for Environmental Economics. She has extensive expertise in exposure assessment, environmental epidemiology, risk assessment, and research translation. Her research interests are in evaluating prenatal exposures to environmental chemicals and related adverse pregnancy outcomes and characterizing developmental risks. She has also been involved in developing guidance for applying evidence-based medicine methodology to environmental health sciences. Dr. Woodruff was appointed to the California Developmental and Reproductive Toxicant Identification Committee in 2012. She received her PhD in bioengineering from a joint program of the University of California, Berkeley and San Francisco.

Lauren Zeise is deputy director for scientific affairs in the California Environmental Protection Agency's Office of Environmental Health Hazard Assessment. She oversees the department's scientific activities, which include the development of risk assessments, hazard evaluations, toxicity reviews, cumulative impacts analyses, and frameworks and methods for assessing toxicity and cumulative impact and the department's activities in the California Environmental Contaminant Biomonitoring Program. Dr. Zeise was the 2008 recipient of the Society of Risk Analysis's Outstanding Practitioner Award. She has served on several advisory boards and committees of the US Environmental Protection Agency, the Office of Technology Assessment, the World Health Organization, and the National Institute of Environmental Health Sciences. Dr. Zeise has served on numerous National Research Council and Institute of Medicine committees, including service as a member of the Committee on Improving Risk Analysis Approaches Used by the US Environmental Protection Agency and current service on the Committee on the Review of the Formaldehyde Assessment in the National Toxicology Program 12th Report on Carcinogens. She received her PhD from Harvard University.

Yiliang Zhu is professor in the Department of Epidemiology and Biostatistics of the University of South Florida, where he directs the Center for Collaborative Research. He is also professor of internal medicine at Morsani College of Medicine. His current research involves quantitative methods in health risk assessment, including physiologically based pharmacokinetic models, dose–response modeling, benchmark-dose methods, and uncertainty quantification. He also conducts research in health outcome and health system evaluation. Dr. Zhu is currently a Fulbright Research Fellow in China, where he is establishing a 15-year cohort study on rural health and human development in northwest China. Dr. Zhu has served on several National Research Council committees, including current service on the Committee to Review the IRIS Process. He received his PhD in statistics from the University of Toronto.