UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

October 30, 2013

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held July 30 - August 1, 2013 on Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening

TO: Steven Bradbury, Ph.D.
Director
Office of Pesticide Programs

David Dix, Ph.D.
Director
Office of Science Coordination and Policy

FROM: Joseph E. Bailey
Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy
Joseph E Bailey 10/30/13

THRU: Laura Bailey
Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy
Laura Bailey 10/30/13

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, VA on July 30 - August 1, 2013. This report addresses a set of scientific issues associated with Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening.

Enclosure

cc:

Jim Jones
Louise Wise
William Jordan
Margie Fehrenbach
Donald Brady
Jack Housenger
Susan Lewis
Lois Rossi
Robert McNally
Jay Ellenberger
Richard Keigwin
Mark Hartman
Jess Rowland
Thomas Steeger
Greg Akerman
John Liccione
Catherine Aubee
Amy Blankinship
Mary Manibusan
Steven Knott
Patience Browne
Dale Kemery
Linda Strauss
Chris Zarba
Molly Hooven
OPP Docket

FIFRA Scientific Advisory Panel Members

Daniel Schlenk, Ph.D.
Barry Delclos, Ph.D.
Marion Ehrich, Ph.D., DABT, ATS
Steven Klaine, Ph.D.
Martha Sandy, Ph.D., M.P.H.

FQPA Science Review Board Members

Robert Denver, Ph.D.
Roberta Grant, Ph.D.
Nelson Horseman, Ph.D.
Jacob Kerby, Ph.D.
Jack Leonard, Ph.D.
Deborah MacLatchy, Ph.D.
Edward Perkins, Ph.D.
Lynnda Reid, Ph.D.
Katherine Roby, Ph.D.
Geoffrey Scott, Ph.D.
Daniel Selvage, Ph.D.
Terry Schultz, Ph.D.

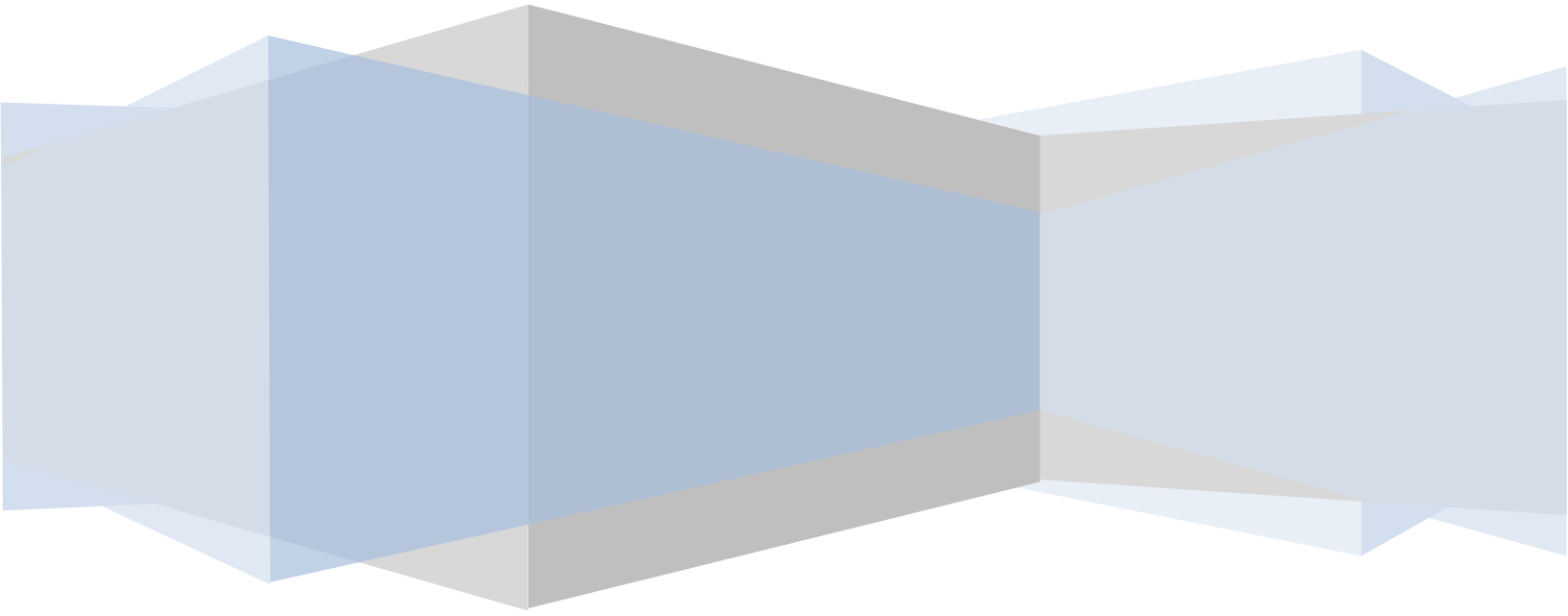
SAP Minutes No: 2013-05

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening

FIFRA Scientific Advisory Panel Meeting
Held at the Environmental Protection Agency
Potomac Yard Conference Center

July 30 – August 1, 2013



NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). The meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of the meeting minutes does not represent information approved or disseminated by the Agency. The meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of these meeting minutes do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal Government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the Environmental Protection Agency, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA Science Review Board members serve the FIFRA SAP on an ad hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Joseph E. Bailey, SAP Designated Federal Official, via e-mail at bailey.joseph@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented by public commenters.

TABLE OF CONTENTS

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening

July 30 – August 1, 2013
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

Daniel Schlenk, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel
Date: OCT 3 0 2013

Joseph E. Bailey
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: OCT 3 0 2013

Federal Insecticide Fungicide and Rodenticide Act
Scientific Advisory Panel Meeting
July 30 – August 1, 2013

PARTICIPANTS

FIFRA SAP Chair

Daniel Schlenk, Ph.D., Professor of Aquatic Ecotoxicology & Environmental Toxicology Department of Environmental Sciences, University of California, Riverside, Riverside, CA

Designated Federal Official

Joseph E. Bailey, FIFRA Scientific Advisory Panel, Office of Science Coordination & Policy, U.S. Environmental Protection Agency,

FIFRA Scientific Advisory Panel Members

K. Barry Delclos, Ph.D., Pharmacologist, Division of Biochemical Toxicology, National Center for Toxicological Research, U.S. Food & Drug Administration, Jefferson, AR

Marion Ehrich, Ph.D., DABT., ATS, Professor of Pharmacology and Toxicology, Department of Biomedical Science and Pathobiology, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, VA

Steven Klaine, Ph.D., Professor and Director, Clemson University Institute of Environmental Toxicology, Department of Biological Sciences, Pendleton, SC

Martha Sandy, Ph.D., M.P.H., Supervising Toxicologist and Chief, Reproductive and Cancer Hazard Assessment Branch, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, CA

FQPA Science Review Board Members

Robert Denver, Ph.D., Professor, Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI

Roberta Grant, Ph.D., Manager, Toxicology Section, Toxicology Division, Office of the Executive Director, Texas Commission on Environmental Quality, Austin, TX

Nelson Horseman, Ph.D., Professor, Departments of Molecular and Cellular Physiology, and Medicine, University of Cincinnati, Cincinnati, OH

Jacob Kerby, Ph.D., Director of Research, Missouri River Institute, Assistant Professor, Biology Department, University of South Dakota, Vermillion, SD

Jack Leonard, Ph.D., Professor, Department of Microbiological and Physiological Systems, University of Massachusetts Medical School, Worcester, MA

Deborah MacLatchy, Ph.D., Vice-President, Academic and Provost, Professor of Biology Wilfrid Laurier University, Waterloo, Ontario, Canada

Edward Perkins, Ph.D., Senior Research Scientist, Environmental Networks and Genetic Toxicology, Environmental Laboratory, U.S. Army Engineer Research and Development Center, U.S. Army Corps of Engineers, Vicksburg, MS

Lynnda Reid, Ph.D., Supervisory Pharmacologist/Toxicologist, Center for Drug Evaluation and Research, Office of New Drugs, US Food and Drug Administration, Silver Spring, MD

Katherine Roby, Ph.D., Research Associate Professor, Department of Anatomy and Cell Biology, Director, Reproductive Endocrinology Laboratory, Women's Health Specialty Center Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, University of Kansas Medical Center, Kansas City, KS

Geoffrey Scott, Ph.D., Director, Center for Coastal Environmental Health and Biomolecular Research, National Centers for Coastal Ocean Science, National Ocean Services, National Oceanic and Atmospheric Administration, Charleston, SC

Daniel Selvage, Ph.D., Associate Professor, College of Pharmacy, University of New England, Portland, ME

Terry Schultz, Ph.D., Emeritus Professor, Department of Comparative Medicine, College of Veterinary Medicine, University of Tennessee, Knoxville, TN

INTRODUCTION

The Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) has completed its review of the scientific issues associated with Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening. Advance notice of the meeting was published in the Federal Register on April 17, 2013. The review was conducted in an open Panel meeting held in Arlington, VA, on July 30 – August 1, 2013. Dr. Daniel Schlenk chaired the meeting. Joseph E. Bailey served as the Designated Federal Official. Opening remarks at the meeting were provided by David Dix, Ph.D., Acting Director, Office of Science Coordination and Policy (OSCP) and Steven Bradbury, Ph.D., Director, Office of Pesticide Programs (OPP). Overview and technical presentations were given by Mary Manibusan and Patience Brown, Ph.D., of OSCP and Thomas Steeger, Ph.D., Gregory Akerman, Ph.D., John Liccione, Ph.D., Amy Blankinship, M.S. and Catherine Aubee, M.P.A., all of OPP.

EPA developed the Endocrine Disruptor Screening Program (EDSP) in response to FFDCA section 408(p) which requires EPA to "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 as the Administrator may designate." 21 U.S.C. 346a(p)(1). In addition, the provision in section 1457 of the Safe Drinking Water Act (SDWA) provides EPA with discretionary authority to provide for testing under the screening program "of any other substance that may be found in sources of drinking water if the Administrator determines that a substantial population may be exposed to such substance." (42 U.S.C. 300j-17).

Based on recommendations from the Endocrine Disrupter Screening and Testing Advisory Committee (EDSTAC) and, pursuant to the EPA Administrator's discretionary authority, the EPA expanded the program to encompass the estrogen, androgen, and thyroid (E, A, and T) hormonal pathways of the endocrine system and human and ecological effects. Subsequent to review by a joint committee of the EPA's Science Advisory Board (SAB) and the FIFRA SAP, the EDSP embarked on a validation process as mandated to evaluate the relevance and reliability of Tier 1 screening and Tier 2 test methods. As recommended by a FIFRA SAP, the current EDSP Tier 1 screening battery consists of both *in vitro* and *in vivo* assays that provide redundancy within a particular mode or pathway of action and complementary endocrine specific-endpoints sensitive enough to detect effects on estrogen, androgen, and thyroid (E, A and T) signaling through different routes of exposure and across multiple life-stages and taxa. The degree of redundancy and complementary assays/endpoints are intended to provide corroborating information to support an evaluation of the Tier 1 screening results.

EPA issued the first test orders of the EDSP Tier 1 screening on 67 chemicals (List 1 chemicals) between October 29, 2009 and February 26, 2010 (*http://www.epa.gov/endo*). As a result of these test orders, EDSP Tier 1 data were submitted on 50 pesticide active ingredients and 2 pesticide inert ingredients. For some test orders, EPA accepted "other scientifically relevant information" (OSRI) in lieu of specific study data (*http://www.epa.gov/endo*).

In May 2013, the Agency held a FIFRA SAP meeting to obtain input to ensure that individual assays and the overall battery performed as anticipated toward understanding whether a chemical is impacting E, A, and T pathways. A subset of the List 1 chemicals were presented to the Panel to evaluate whether each assay can be consistently executed based on the performance criteria and to discuss any issues associated with interpretation of the responses within each assay as well as the anticipated complementary relationships both within and across the assays. The advice and recommendations of the Panel from the May FIFRA SAP were critical in how the Agency conducted its weight-of-evidence (WoE) evaluation of the Tier 1 screening results, which was the topic of this FIFRA SAP.

The EPA issued its WoE guidance document in 2011 for evaluating the results of EDSP Tier 1 screening to identify the need for Tier 2 testing. That document can be found at www.regulations.gov (Docket ID number EPA-HQ-OPPT-2010-0877). Briefly, that document presents a hypothesis-based approach that begins with an evaluation of each study's quality and relevance in addressing the questions for the chemical of interest, and guidance on how to assemble and integrate all lines of evidence (EDSP Tier 1 assays and OSRI, including peer reviewed studies) for that chemical. Thus, Tier 1 screening is combined with other relevant evidence (e.g., CFR part 158 guideline studies) using a WoE analysis intended to determine whether or not a test chemical requires more comprehensive Tier 2 testing or a more targeted and tailored approach.

The Agency presented four case studies based on a subset of List 1 chemicals for the Tier 1 test orders and one additional case study for a chemical that was not subject to the Tier 1 test orders to illustrate the decision logic for applying EPA's EDSP WoE guidance (*http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2010-0877-0021*) in interpreting Tier 1 screening results and OSRI. The FIFRA SAP was asked to comment on interpretative issues that arose during the WoE approach as well as the decision logic that guided the determination of whether higher level testing is needed.

PUBLIC COMMENTS

Oral Statements were presented as follows:

Patricia Bishop, M.S. on behalf of People for the Ethical Treatment of Animals
Catherine Willett, Ph.D., on behalf of the Humane Society of the United States
Scott Slaughter on behalf of the Center for Regulatory Effectiveness
Thomas G. Osimitz, Ph.D., Science Strategies, LLC, on behalf of the Alkylphenols & Ethoxylates Research Council
Ellen Mihaich, Ph.D., DABT, of Environmental & Regulatory Resources, LLC; Chris Borgert, Ph.D. of Applied Pharmacology & Toxicology, Inc.; Pat Kwiatkowski, Ph.D., of Bayer CropScience; Sue Marty, Ph.D., DABT, of The Dow Chemical Company; Barbara Neal, DABT, of Exponent; and Elliot Gordon, Ph.D., DABT, of Makhteshim Agan all on behalf of the Endocrine Policy Forum

Written Statements were provided by:

Anonymous
Richard A. Becker, Ph.D., DABT and Emily V. Tipaldo, M.A. on behalf of the American Chemistry Council
Patricia L. Bishop, M.S. on behalf of People for the Ethical Treatment of Animals
Scott Slaughter on behalf of the Center for Regulatory Effectiveness
Catherine Willett, Ph.D. on behalf of the Humane Society of the United States
Barbara S. Losey on behalf of the Alkylphenols & Ethoxylates Research Council
Clare Thorpe, Ph.D. on behalf of the Endocrine Policy Forum

LIST OF SELECTED ACRONYMS

A	Androgen (hormonal pathway)
AChE	Acetylcholinesterase
AMA	Amphibian Metamorphosis Assay
AOP	Adverse Outcome Pathway
AR	Androgen Receptor
ASTER	ASsessment Tools for the Evaluation of Risk
Bmax	Binding at maximum
CV	Coefficient of Variation
DER	Data Evaluation Record
EDSTAC	Endocrine Disruptor Screening and Testing Advisory Committee
EDSP	Endocrine Disruptor Screening Program
E	Estrogen (hormonal pathway)
ER	Estrogen Receptor
ERTA	Estrogen Receptor Transcriptional Activation
FFDCA	Federal Food, Drug and Cosmetic Act
FIFRA	Federal Insecticide, Fungicide and Rodenticide Act
FQPA	Food Quality Protection Act
FSTRA	Fish Short-Term Reproduction Assay
GSI	Gonado-Somatic Index
HPG	Hypothalamic-Pituitary-Gonadal Axis
HPT	Hypothalamic-Pituitary-Thyroidal Axis
IC50	Inhibitory Concentration at 50% of response
Koc	Soil-water partition coefficient
Kow	Octanol-water partition coefficient
Kd	Equilibrium Dissociation Constant
LOAEL	Lowest observed adverse effect level
MIE	Molecular Initiating Event
MoA	Mode of Action
MTD	Maximum Tolerated Dose
NOAEL	No observed adverse effect level
OCSPP	Office of Chemical Safety Pollution and Prevention
OECD	Organization for Economic Co-Operation and Development
OP	Organophosphorus
ORD	Office of Research and Development
OSCP	Office of Science Coordination and Policy
OSRI	Other Scientifically Relevant Information
PPS	Preputial seperation
SAB	Science Advisory Board
SAP	Scientific Advisory Panel
SDWA	Safe Drinking Water Act
T	Thyroid (hormonal pathway)
T4	Thyroxine (tetraiodothyronine)
TBG	Thyroxine-binding globulin

ToxCast	A chemical library that consists of HTP data generated by EPA and a series of contract laboratories.
TP	Testosterone Propionate
TR	Thyroid Receptor
TSH	Thyroid Stimulating Hormone
VO	Vaginal opening
VTG	Vitellogenin
WoE	Weight-of-evidence

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Charge Issue 1.1 - Please comment on whether the Agency has transparently described the conduct and results of the individual Tier 1 studies and the OSRI for each of the case studies (Sections 6-9 of the white paper), and specifically whether the level of detail is sufficient to ensure that a study is reliable for determining the potential to interact with E, A, or T signaling pathways and the rationale for the preliminary study conclusion. The Panel concluded that in general, the Agency described the conduct and results of the Tier 1 studies and OSRI in a highly transparent manner in the case studies for each of the five Chemicals A, S, J, N and X. In general, the level of detail provided in the White Paper for each case study was sufficient to evaluate whether a study is reliable or not for determining the potential to interact with the E, A or T signaling pathways. The level of detail in the discussion and conclusions sections of the case studies was not always sufficient to determine the rationale for the preliminary study conclusions. The Panel found that the current review process: 1) individual assay review (by contractor and EPA), 2) Tier 1 EPA technical review of groups of related assays, and 3) WoE technical review of all Tier 1 data and OSRI by an EPA expert panel is a sound process that ensures thoroughness, impartiality and completeness of the review process. The format of the case studies presents the relevant information in a clear and logical manner and with sufficient detail to evaluate applicability and utility. The Panel did indicate that while completeness of Tier 1 data may be easily assessed, it may be difficult to ascertain completeness of OSRI. The Panel found the supplemental case study for Chemical X particularly intriguing as it was constructed exclusively from OSRI. Some Panel members suggested that such OSRI studies may be useful in prioritizing the need for individual Tier 1 studies, especially when there is a large amount of good quality data available from OSRI sources. Some Panel members thought that as the process moves forward, there are some possibilities for quantification of the review process (as opposed to the data *per se*). For example, as more WoE analyses are completed it may be possible to develop a relative ranking for all the Tier 1 endpoints for each of the eight E, A or T pathway-related outcomes. In addition, some Panel members thought that it may be possible to capture the degree of consistency of opinions expressed by the WoE reviewers and that would enhance the transparency of the process and help better address the uncertainty associated with each determination, particularly as data are considered for inclusion or exclusion in the final decision-making process for each chemical. While only five chemicals are reported, the current framework does not appear to address how the WoE process will use OSRI data that falls outside of results for the E, A and T assessments. How this type of data will be used in the final decision for a Tier 1 assessment is unclear and may become important, especially if more data are used from other existing databases.

Charge Issue 1.2 - For each of the case studies, please comment on whether the performance criteria are clearly stated for the Tier 1 assays and, when results were not within the boundaries of the performance criteria, whether EPA has clearly expressed why the data are still considered reliable. In general, the discussion of performance criteria in the White Paper varied across assays and case studies, with more detail on specifics typically given for the *in vitro* assays. The specifics of the criteria for each assay were often not stated. When criteria were not specifically stated, it was assumed they were met, although in some cases it was explicitly stated that all criteria were met. The description of the rationale for accepting data that

were outside the performance criteria also varied, although often it was clear that the discrepancies were minor and occurred in only one or a few endpoints so that they did not affect interpretation of the assay as a whole. Statements such as "performance criteria were generally met" or "performance was reasonable" often appear in the White Paper, but are not particularly descriptive. A consistent format for the statements of the criteria for each assay, whether the assays met the criteria, and why this was or was not a concern would be useful. The detailed response to Charge 1.2 provides what the Panel found were specific inconsistencies across compounds in the manner in which the criteria were presented and in which the rationale for the acceptance of data that did not meet the criteria was described.

Charge Issue 1.3 - Please comment on the Agency's conclusion regarding the utility of the AMA data for Chemical S to still reliably evaluate its potential endocrine interaction in a WoE analysis. The concentrations of Chemical S achieved in the aquarium water in the AMA were below the targeted nominal concentrations, and there was large variability in the measured concentrations. For these reasons the Panel disagreed with the Agency's conclusion that the exposure was reasonably quantified. Because of the low solubility and high variability in the measured concentrations of Chemical S, and a lack of information on how the measured concentrations of Chemical S compare to expected environmental exposures, the Panel concluded that the results do not enable reviewers to determine responses.

Despite the difficulty in interpreting findings with Chemical S, the Panel agreed with the Agency's conclusion that the assay should not be repeated using the same route of administration. However, the Panel recommended that the Agency develop protocols for the AMA (and the Fish Short-term Reproduction Assay (FSTRA)) to assess problematic chemicals like Chemical S that have low aqueous solubility. If the goal of the AMA is to evaluate ecotoxicological potential, then the chemicals should be tested at concentrations, and by routes that are ecologically relevant. The Panel considered that the limited data do not allow one to conclude that Chemical S "does not demonstrate a potential to interact with the thyroid pathway".

Charge Issue 2.1a - Please comment on how the Agency applied its decision logic to integrate an understanding of overt toxicity in the context of observed Tier 1 in vivo responses, and in particular, the Agency's determination not to place weight on the FSTRA high concentration responses coincident with overt toxicity. The Panel commended the Agency for including Figure 3, FSTRA decision logic diagram, because it provides very useful information on the diagnostic utility of the different endpoints and ranks them from greatest to least significance. The Panel recommended that the Agency provide similar decision logic algorithms for the other Tier I assays, especially when more than one endpoint is evaluated within an assay. The Agency has successfully applied the correct decision logic for Chemical A based on dose selection and diagnostic utility for assessment of endocrine activity except for female dosing. The Panel recommended a more detailed discussion be included regarding this endpoint in the WoE discussion for Chemical A. Based on the WoE guidance, the Agency has concluded that at the highest concentration, the observed endocrine effects were likely due to a non-endocrine mode of action (MoA) (i.e., cholinergic intoxication) and general toxicity; and, has discounted the endocrine effects that occur when overt toxicity occurs. The Panel agreed with this finding. In addition, the Agency also noted the directionality of the FSTRA responses supported this

decision (i.e., decreases in the measured endpoints reflected a compromised organism with limited ability to maintain reproductive function and homeostasis). The Panel also agreed with this finding.

Charge Issue 2.1b - Please comment on how the Agency applied its decision logic to integrate an understanding of overt toxicity in the context of observed Tier 1 in vivo responses, and in particular, on the Agency's determination to place less weight on the Tier 1 in vivo responses in the presence of overt toxicity. The Panel accepted the Agency's interpretation of overt toxicity data as insufficient information to conclude endocrine disruption. There are clear issues with the overlap of alternate pathways to adequately evaluate endocrine disruption. The current methodology of Tier 1 testing does not allow for adequate interpretation of endocrine disruption at these overtly toxic levels. Addressing Chemical S in particular, there are clear issues of solubility. That being stated, each of the resulting effects found at the high concentration are consistent with the proposed toxic MoA (decreases in testosterone, decreases in male and female gonadal weights, delays in vaginal opening (VO) and preputial seperation (PPS), decreases in male fertility, and increases in male gonado-somatic index (GSI) and vitellogenin (VTG). The Panel found that these results were potentially the result of a toxic level of exposure and, therefore, did not provide sufficient evidence for endocrine disruption on their own. The Panel agreed that other findings (androgen-dependent tissue weights in Hershberger assay), at levels that do not show overt toxicity, do show the potential for interaction in the androgen pathway. While there was some question concerning interaction with the estrogen pathway, the Panel agreed with the Agency's overall assessment of Chemical S.

Charge Issue 2.1c - Please comment on the Agency's analysis in characterizing Tier 1 responses that are expressed at or near limit doses where some degree of overt toxicity occurs, and the extent to which such responses are considered in the WoE analysis. Chemical N is a good example of an agent that had effects at toxic doses, as anti-androgenic and estrogenic actions occurred in the FSTRA, but with one exception, this was only at overtly toxic doses. The Panel agreed with the Agency's approach to put lowered weight on results from assays where overt toxicity occurred. In general, the Panel found that the Agency should reduce the potential for high dose effects by not testing at or near dose-limits as determined in the range finding studies. This could be done by more testing at lower doses; in particular to the FSTRA, it was noted by a Panel member that due to the logarithmic dosing paradigm, the Agency has ample room to try a range of lower doses without adding a significant amount of time to run the assay. Overall, the Panel concluded that assays exhibiting overt toxicity are not useful for interpretation of whether a compound has an endocrine effect, and that results for assays using doses near the test limits need to be evaluated carefully.

Charge Issue 2.1d - Please comment on the Agency's overall approach to characterizing Tier 1 responses coincident with overt toxicity and determining the weight to be given to such responses. The Panel and EPA are in agreement that overt toxicity is a confounder and the Panel noted that even more consideration should be taken in evaluating Tier 1 endocrine endpoints in relationship to non-endocrine toxicity and when selecting the highest dose/concentration to be used in Tier 1. Evidence of endocrine toxicity in the presence of compound-induced toxicity should be highly scrutinized and suspect. Dose range finding studies or some other dose/concentration titration paradigm should be used to determine the exposure/dose-limiting

toxicity to avoid confounding the interpretation of endocrine related effects at overtly toxic exposure levels. The Panel recommended that EPA standardize the criteria used for determining overt toxicity that confounds interpretation of data. In summary, the Panel agreed that little, if any, weight should be placed on signs of endocrine disruption in the presence of overt toxicity. All effects in endocrine sensitive tissues should be evaluated in terms of primary interactions with the endocrine system vs. secondary effects related to toxicity in non-endocrine organs or overall disruptions in homeostasis.

Charge Issue 2.2 - For Chemicals N and A, please comment on the decision logic the Agency has used to characterize these types of situations where there is a lack of robustness in terms of complementarity and redundancy, and the transparency and reasonableness of the approach. With regard to Chemical N, most Panel members generally agreed that EPA's findings followed the decision logic within the White Paper and EPA's conclusions were supported by OSRI data and other outside references that were peer reviewed and quality assured for inclusion in this process. For Chemical A, the Panel thought that the OSRI data clearly supported EPA's conclusion to exclude the FSTRA effects as supporting an estrogen-related effect. This lack of complementarity and redundancy in the *in vivo* assays was affected by the confounding nature of overt toxicity for Chemical A, which led EPA to conclude via a WoE approach that there was a lack of consistency within these parts of the Tier 1 battery.

Finding unambiguous effects in one or more *in vivo* assays, in the absence of any overt toxicity, might be considered sufficiently powerful evidence to avoid consideration of additional complementary or redundant evidence. In particular, these diagnostic *in vivo* responses explicitly do not rest on any *in vitro* or mechanistic evidence (receptor binding, transcriptional assays, etc.). For example, if there is a clear uterotropic effect, no amount of negative evidence in the estrogen receptor (ER) binding, steroidogenesis, or aromatase data is sufficient to negate this finding simply because of lack of coherence or complementarity. The Panel believed it is reasonable for EPA to consider adopting or developing ranked or tiered approaches, such as those presented during public comments, to better define the WoE evaluation process and move toward one that is more quantitative.

Charge Issue 2.3 - Please comment on the how the Agency has characterized the endocrine interaction for Chemical J at different levels of biological organization across taxa, and the transparency and reasonableness of the conclusions drawn. Please include in your response, comments regarding the Agency's conclusion about differences in sensitivities between taxa (i.e., fish and rats), regarding chemicals that appear to alter steroidogenesis. The Panel concurs with the Agency that Chemical J does interact with the estrogen signaling pathway through effects on steroidogenesis. The conclusion, from a WoE approach, that there are effects on steroidogenesis is supported through redundancy (in the *in vitro* steroidogenesis and aromatase assays and the *in vivo* FSTRA) and complementarity (endpoints within the FSTRA). Lack of redundancy between the *in vivo* female rat pubertal assay and the *in vivo* FSTRA could be due to differences in taxon sensitivity. Panel members noted that the difference in female fish body weight in the FSTRA at the highest Chemical J concentration indicates a potential confounding factor that should be investigated further in the FSTRA protocol.

Charge Issue 2.4 - Please comment on how the Agency has integrated different sources of data along a biological continuum to characterize endocrine interactions of Chemical A and the transparency and reasonableness of the decision logic. The case study for Chemical A clearly raises the problem that the battery of assays may yield results that are largely unsupported by other assay battery findings. In such cases, what would be the logic for making further recommendations? Since Tier 1 is a "screening" exercise, an argument can be made that any single reliable positive finding, which the Hershberger antagonist findings appear to be, would be sufficient to move Chemical A to some next level of testing. However, an equally powerful case could be made that a single test, such as one arm of the Hershberger, is insufficient evidence to consider Tier 2 testing without having either a reliable mechanism of action to explain it, or other complementary evidence in different *in vivo* assays. It remains unclear how these choices will be made. In order for a molecular initiating event (MIE), such as receptor binding, to be relevant as an endocrine disruptor that drives a particular adverse outcome pathway (AOP) will need certain quantitative relationships established. In the context of a WoE discussion, characterization of a compound as a "binder" does not provide sufficient information to determine with any precision how much weight should be placed on this particular evidence. Since the primary pesticidal MoA for Chemical A is cholinesterase inhibition, and the compound does not appear to accumulate, one assumes that it would manifest toxicity via cholinesterase inhibition at concentrations much below those implied for androgen pathway activity by the extremely weak androgen receptor (AR) affinity. With sufficient information this could be calculated, and the Panel recommended this approach.

Charge Issue 2.5 - For Chemicals N and S, please comment on the how the Agency has integrated different sources of data along a biological continuum to characterize the endocrine interaction and the transparency and reasonableness of the conclusion drawn. The Panel agreed that negative Tier 1 *in vitro* data do not detract from or limit the conclusion of a potential endocrine interaction based on positive data observed in the *in vivo* assays only. As indicated by the Agency, it is entirely possible that an unanticipated molecular initiating event or metabolism of the chemical to an active intermediate may account for interaction with the endocrine system. Based on the case studies presented in the White Paper for Chemicals N and S, the Agency has included a large body of OSRI representing a range of organisms and tests in the WoE analysis. Further, the Agency has integrated this information into the overall WoE analysis for Chemical N and S. The Panel found the information proved to be transparent and the conclusions drawn to be reasonable.

Charge Issue 2.6 - The Agency considered the lack of complementarity and redundancy in responses to support a conclusion of no interaction with the HPT axis, and viewed these isolated responses insufficient to support a conclusion of an interaction with the thyroid signaling pathway. Please comment on the how the Agency has characterized the endocrine interaction at different levels of biological organization, and the transparency and reasonableness of the conclusion drawn. Based on the limited repertoire of *in vivo* hypothatlamic-pituitary-thyroid (HPT) axis assays in the Tier 1 battery, the Agency is correct in that there is insufficient data to conclude that any of the test compounds interact with the thyroid-signaling pathway. The current Tier 1 assays for the HPT axis cannot provide the degree of confidence for screening that is available for the E and A axis. For screening purposes, thyroid

stimulating hormone (TSH) and thyroxine (T4) measurements are used by clinicians worldwide to assess thyroid status and remain the gold standard in vertebrate studies.

Charge Issue 2.7 - In the absence of Tier 1 data, OSRI was available for Chemical X that indicated effects on thyroid endpoints in the rat but the results were inconsistent within and among studies and there was no OSRI presented from amphibian studies. Because of studies that were not specifically validated to detect an interaction with the thyroid hormonal pathway, limited data, and ambiguous results, the potential for Chemical X to interact with the thyroid pathway cannot be excluded. Please comment on the how the Agency has characterized this endocrine interaction at different levels of biological organization, and the transparency and reasonableness of the conclusion drawn. The OSRI provided for Chemical X is not sufficient to exclude the HPT axis as a potential target and further study is required. Similar to the discussion in Charge 2.6, the multiple sites for potential xenobiotic interactions along the HPT axis require additional studies to adequately evaluate interactions between Chemical X and the HPT axis. Based on the limited OSRI dataset, the Panel concluded the Agency appropriately evaluated the potential for Chemical X to impact the HPT axis.

Charge Issue 3.0 - Based on all of the case study analyses, please provide overall comments on how the Agency has employed its WoE guidance and characterized the evidence and conclusions and include in your response the following points:

a. How consistent and transparent the cases studies are in terms of documentation. In general, the case studies are consistent and transparent; however, there were problems in applying overt toxicity designations, the rationale for not discarding data where controls did not meet guidelines, and how different endpoints and assays were weighted. Transparency would be greatly facilitated if mode of action for related and characterized chemicals were also considered. There was consensus on the Panel that the overall approach need not necessarily be highly quantitative for a screening level set of tests, but the use of a systematic and transparent decision tree would be of great use to everyone. Without the presence of the decision trees, which might be difficult to create for all the Tier 1 tests, a clear description of how and why the decisions were made to indicate the potential (or lack of potential) for interaction along the pathways should be included in this integrative fashion for the discussion of chemicals.

b. How adequately the Agency has described the extent of complementarity and redundancy of responses and has integrated and interpreted diverse lines of evidence across different biological levels of organization and taxa to reach preliminary conclusions regarding endocrine interactions. More visualization, including converging AOPs and decision logic trees, will improve understanding and transparency. The WoE process should move towards a more quantitative approach integrating high throughput data and computational approaches. Overall, the Panel thought the Agency adequately described the extent of complementarity and redundancy of responses. The Panel pointed this out for Chemical J, where there are reasonable scientific reasons to accept lack of complementarity and redundancy. The Agency should provide guidance on how potential species-specific responses should be assessed.

c. How the Agency has used OSRI data to further characterize the observations from EDSP Tier 1 assays in determining potential chemical interactions with the E, A, and T signaling pathways. The Panel thought OSRI was helpful in discerning overt toxicity for Chemical A. The current framework does not address how it will use OSRI data that falls outside of current E, A and T assessments in the current Tier 1 Screen as observed with invertebrate data available for Chemical J. A more defined process for adding or removing data for inclusion/exclusion should be developed with OSRI. Toxicological data from related chemicals, high throughput data, and computational models could provide more support for E, A and T pathway impacts, especially where effects in one species are not supported by corresponding effects in another.

d. Chemical mode of action and weight placed on Tier 1 responses in the presence of uncertainties introduced by dose setting, overt toxicity, and portal of entry issues. ASTER (ASsessment Tools for the Evaluation of Risk) and other toxicity databases should be fully utilized as a Tier 1 pre-screen for as many chemicals as possible. The case study for Chemical X based on OSRI only was a very good indicator of the utility of this approach. The log-based dose design of some assays limit the ability of assays to produce dose response relationships. Selection of doses or addition of more doses should be considered to avoid overt toxicity. Exposure route (portal of entry) may be a critical consideration. Dimensionality and magnitude of effects help determine the certainty and consistency of results leading to robustness, which EPA demonstrated with several examples in the case studies. In cases where overt toxicity is observed, the Agency should not make conclusions on endocrine potential.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

The purpose of the current review is to seek comment from the FIFRA Scientific Advisory Panel (SAP) on the implementation of the basic interpretive process, principles and concepts outlined in the Agency's 2011 EDSP Weight-of-Evidence (WoE) guidance (Appendix 1, see Section 3 of the White Paper) as illustrated with case studies.

The current case study analyses are primarily focused on the interpretative process for determining whether or not there is a chemical interaction with estrogen, androgen, or thyroid (E, A, or T) signaling pathways. An important aspect of addressing this question includes characterizing at what level of biological organization (e.g., molecular, cellular, tissue/organ, organism) that interaction is expressed, the robustness of the data showing an interaction in terms of the extent of complementarity and redundancy across multiple lines of evidence, and describing under what conditions (e.g., in an organism with an intact hypothalamic-pituitary-gonadal axis (HPG), in the absence of overt toxicity, at what doses, duration and route of exposure, etc.) the chemical interacts with and perturbs the endocrine system.

The charge questions are organized around the fundamental steps in the WoE guidance (Section 3 of the White Paper) in addressing whether or not a chemical interacts with the E, A, or T signaling pathways: (i) assembling and evaluating the scientific quality of the individual studies, and (ii) formulating hypotheses and integrating data at different levels of biological organization, which includes characterizing the extent and nature of the complementarity and redundancy in responses. In the context of these steps in the WoE guidance, cross-cutting as well as case study-specific questions are provided, consistent with the rationale for selecting the example chemicals for this review (see Section 4, Table 2 in the White Paper).

As described in the 2011 EDSP WoE guidance, in assembling and evaluating the quality of scientific information to support a determination of a chemical's potential to interact with E, A, or T signaling pathways, it is important to ensure data soundness, applicability, utility, clarity and completeness.

Charge 1.1. Please comment on whether the Agency has transparently described the conduct and results of the individual Tier 1 studies and the OSRI for each of the case studies (Sections 6-9 of the White Paper), and specifically whether the level of detail is sufficient to ensure that a study is reliable for determining the potential to interact with E, A, or T signaling pathways and the rationale for the preliminary study conclusion.

Panel Response

As noted by the United States Environmental Protection Agency (the Agency), the purpose of the current review is to seek comment from the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) on the implementation of the basic interpretive process, principles, and concepts outlined in the Agency's 2011 Endocrine Disruptor Screening Program (EDSP) Weight-of-Evidence (WoE) guidance which is presented in

Appendix 1 of the Agency's White Paper and illustrated with case studies presented in Sections 6 – 9 of the White Paper and the supplemental case study, Chemical X.

The Panel understood that this SAP review focused on providing comments on how the Agency is employing the 2011 EDSP WoE guidance on integrating evidence from the Tier 1 battery and other scientifically relevant information (OSRI) beyond Tier 1 data to determine whether or not a test chemical interacts with the estrogen, androgen or thyroid signaling pathways. The Panel further understood it was not the mission of this SAP to comment on the EDSP WoE guidance itself, since the guidance is based on longstanding practices and principles articulated in other EPA peer-reviewed documents.

As noted in the White Paper (Section 3) the cornerstone of the Agency's 2011 EDSP WoE guidance is a hypothesis-based approach following the adverse outcome pathway (AOP) and mode of action (MoA) conceptual frameworks. Briefly, these frameworks integrate data along a biological continuum of key events from chemical structure through molecular initiating events, responses at the cellular and organ/tissue level, to *in vivo* outcomes. The Panel understood that the WoE process basically addresses one question: Does the chemical have the "potential" to interact with the endocrine system via the E, A, or T signaling pathway? It does this by; 1) assembling the relevant data, 2) evaluating data for quality and relevance, and 3) integrating the different lines of evidence to support conclusions concerning whether a chemical has the potential to interact with the endocrine system.

As described in the 2011 EDSP WoE guidance, in assembling and evaluating the quality of scientific information to support a determination of a chemical's "potential" to interact with E, A, or T signaling pathways, it is important to ensure data soundness, applicability, utility, clarity and completeness. The Panel understood the mode of action framework, particularly the adaption of the Bradford Hill Criteria (e.g., sequence of key events, coherence, strength and limitations, consistency and biological plausibility) is central to evaluating the quality of scientific information.

Crucial to the current SAP are the four case studies presented in Sections 6 - 9 of the White Paper (Chemicals A, S, J and N), which come from the 21 chemicals used for the May 2013 SAP review of Tier 1 Screening Assays and Battery Performance, and the additional case study (Chemical X) provided in the supplemental material. As stated in the White Paper, the aim of the case studies is to illustrate the basic analytical process, principles and concepts outlined in the 2011 EDSP WoE guidance. Because these case studies are illustrative, the Panel understood the analyses presented in this SAP are not to be interpreted as final or complete.

The Panel congratulated the Agency on providing robust case studies with which they have illustrated their use of the WoE guidance. The Panel also thanked the public commenters for their input. Clearly, each case study sparked a number of probing questions.

Only 21 of the 52 chemical for which there is Tier 1 data were evaluated in the May 2013 SAP and, thus, considered for inclusion as case studies in the present SAP; from these 21 chemicals only four were presented as case studies, with an additional case study consisting of OSRI data presented on a chemical that had not undergone Tier 1 testing (Chemical X). The

Agency notes that the five case studies (i.e., Chemicals A, S, J, N and X) presented in the White Paper and supplemental document are illustrative of the likely range of typical responses (e.g., different mode(s) of action, different chemical properties) that will be observed across the Tier 1 assays and OSRI submitted in response to EDSP Tier 1 test orders. Based on the specific information presented for each case study, the Panel agreed with the Agency that the five case studies present a range of responses in the Tier I test battery.

The limited number of case studies has not affected the Panel's ability to comment on the Agency's employment of the 2011 EDSP WoE guidance on integrating evidence from the Tier 1 battery and OSRI to determine whether or not these five chemicals have the potential to interact with the E, A, or T signaling pathways. However, the limited number of case studies did adversely impact the Panel's ability to suggest how lack of complementarity within an assay and redundancy across assays (i.e., inconsistencies) may be looked at in the context of other modes of toxic action, in particular how a given pesticidal mode of action may impact complementarity and redundancy in the Tier I battery. For example, do other organophosphate cholinesterase inhibitors affect the androgen AOP in a pattern similar to Chemical A? The Panel predicted that as the number of chemicals that have undergone a Tier I WoE evaluation increases, the Agency will be able to identify common response patterns associated with certain non-endocrine modes of action (e.g., cholinesterase inhibition, narcosis) and other common response patterns associated with E, A or T modes of action. This information will enable the Agency to better discriminate between truly inconsistent responses in the Tier 1 battery versus responses that are associated with non-endocrine toxicity modes of action, or with newly recognized interactions with the E, A or T signaling pathways.

While the Panel desired to see more case studies, the Panel noted the justification for the case studies presented in the White Paper (see Section 4) is compelling. Since the results of the May 2013 SAP on the review of Tier 1 Screening Assays and Battery Performance was not available at the time of this SAP, the method for interpretation of the endpoints for each of the 11 Tier 1 assays as discussed in Section 5 of the White Paper was positively received by the Panel. The Panel also found the presentation of the decision logic for the amphibian metamorphosis assay (AMA) and the fish short-term reproductive assay (FSTRA) in the White Paper (Figures 2 and 3, respectively) useful in understanding the Agency's interpretation of data generated in these assay systems.

Based on the information presented in the White Paper, the Panel concluded that, in general, the Agency described in a highly transparent manner the conduct and results of the Tier 1 studies and OSRI in the case studies for each of the five Chemicals; A, S, J, N and X. Transparency in using the WoE approach was defined by EPA in their initial presentation as a "descriptive rationale of whether there is sufficient demonstration of the potential to adversely interact with endocrine pathways (E, A and/or T)". Qualification of the consensus of the professional judgment expressed along the different review panels employed by EPA in developing this WoE approach is what is currently reported in the White Paper and case studies presented by EPA. The Panel believed that the current review process: 1) individual assay review (by contractor and EPA), 2) Tier 1 EPA technical review of groups of related assays, and 3) WoE technical review of all Tier 1 data and OSRI by an EPA expert panel is a sound process that ensures thoroughness, impartiality and completeness of the review process.

Furthermore, the Panel believed that the current review process, which focuses on the completeness and consistency of information in regards to endocrine-related MoA/AOP pathways provides important mechanistic relevance between level of biological organization as well as other axes of importance (e.g., across taxa).

The Panel thought that the format of the case studies presents the relevant information in a clear and logical manner and with sufficient detail to evaluate applicability and utility. The Panel did indicate though, that while the completeness of the Tier 1 data is more easily assessed, it might be difficult to ascertain the completeness of OSRI.

In particular, the Panel thought the format (A. Introduction; B. Data Available for the Chemical; C. Tier 1 Screening Assays for the Chemical; D. Other Scientifically Relevant Information (OSRI) for the Chemical; E. Discussion, and F. Conclusion, with subsections to Section C for each of the 11 tier 1 assays; subsections to Section D for *in vitro*, mammalian and ecotoxicity; and subsections to Section E for the estrogen, androgen and thyroid pathways) was particularly useful in their evaluations. Credit is given to the Agency for the summary tables of the lines of evidence indicating potential interaction with each pathway. Following the format noted above allowed the Panel (or a third party) to rather easily evaluate the reliability of the information from each study for the purpose of determining the potential of a chemical to interact with E, A, or T signaling pathways. This format also allowed the Panel (or a third party) to easily evaluate whether the data as described by the Agency supports the given rationale for the preliminary study conclusion.

The Panel recommended that Section C include the following two critical pieces of information: 1) individual assay performance from which the data were collected and 2) assessment of overt toxicity (which may require information to be included that are not measured directly by the assay performance criteria). Conclusions around potential for endocrine interaction are strengthened by the inclusion of these types of information as warranted.

In general, the level of detail provided in the White Paper for each case study was sufficient to evaluate whether a study is reliable or not for determining the potential to interact with the E, A or T signaling pathways.

The level of detail in the discussion and conclusions sections of the case studies was not always sufficient to determine the rationale for the preliminary study conclusions. For example, in the evaluation of the androgen receptor (AR) binding assays with Chemicals A and J, the data indicate that Chemical J has a greater binding affinity than Chemical A, yet the Agency concluded that Chemical A was acting as an AR binder and Chemical J was not. In addition, it was not always clear how uncertainties in the data were taken into account in reaching preliminary study conclusions, as for example, in cases where dose spacing issues (log versus half-log) resulted in the observed effect seen at a dose one log below the next highest (and overtly toxic) dose falling short of a “positive” response (as specified by the decision logic criteria presented in Table 3) (i.e., displacement of 34% vs. 50% of the radio-ligand in a binding assay).

The Panel found the supplemental case study for Chemical X particularly intriguing as it was constructed exclusively from OSRI. Some Panel members believed that such OSRI studies might be useful in prioritizing the need for individual Tier 1 studies, especially when there is a large amount of good quality data available from OSRI sources. Using OSRI studies only fits with the Agency's approach that no one study or endpoint is expected to support a regulatory decision; however, there is a need to ensure that the totality of good, available data have been utilized in this approach.

Some Panel members found the approach used for Chemical X, where there were repeated independent results reported for the same or highly similar endpoint, may make it possible to provide a degree of quantification of the consistency of an endpoint (e.g., uterotrophic assay – uterine weight was increased in 7 out of 7 studies), which is moving this process toward a more quantitative approach. The Panel thought that as more compounds are assessed within the same AOP, it might be possible to include similar measures of consistency for important determinant endpoints using this WoE approach in the future.

The Panel positively received the information that the Agency will consider the collective 2013 SAP reports in evaluating the Tier 1 data for the 52 chemicals in its final WoE determinations for the List 1 chemicals. The Panel thought as the number of WoE exercises increases the Agency would be able to elucidate the inconsistencies (i.e., lack of complementarity within an assay and/or redundancy across assays) in their WoE-AOP evaluations and improve the confidence in the conclusion.

The Panel found that the application of WoE analysis is an integrative and interpretive process that takes into account all relevant scientific information, especially for key events along an AOP. It typically calls for consideration of the existing data, both Tier 1 data and OSRI, in the context of providing relevant, robust, and consistent evidence. In other words, it examines agreement among the outcomes within an individual assay (i.e., complementarity), agreement between different assays or studies within a key event (i.e., redundancy) and, as seen with the OSRI, agreement between outcomes for individual assays representing different key events (i.e., concordance). Moreover, if the data indicate a potential of the chemical to interact with an AOP, it is also important to know at what level(s) of biological organization (e.g., receptor/target binding, cell responses, enzyme activity, tissue/organ response, etc.) interactions are expressed. The impacts of dose, duration of dosing, and route of exposure, as well as metabolism of the compound, in relationship to the modes of toxic action (e.g., narcosis, cholinesterase inhibition, etc) are also critical.

The Panel believed that the application of the WoE concept, as applied by the Agency in the five case studies, shows it to be an integrative and interpretive process evaluating both health and ecological information. While details were not presented, it is a qualitative process (e.g., equivocal, negative, positive) aimed at describing the degree of consistency within the Tier 1 screening data and OSRI for a particular compound. While a conclusion is part of the WoE process, no decision or rational for the decision are included at this time so it is not clear how the Agency will use the WoE outcomes to direct possible further testing.

Some Panel members thought that as the process moves forward, there are some possibilities for quantification of the review process (as opposed to the data *per se*). For example, as more WoE analyses are completed it may be possible to develop a relative ranking for all the Tier 1 endpoints for each of the eight E, A or T pathway-related outcomes.

In addition, some Panel members thought that it may be possible to capture the degree of consistency of opinions expressed by the WoE reviewers and that would enhance the transparency of the process and help better address the uncertainty associated with each determination, particularly as data are considered for inclusion or exclusion in the final decision-making process for each chemical. Furthermore, some Panel members believed that adding information where there was unanimous (100% agreement), majority (>50% < 100% agreement) and divergent (< 50%) consensus would, in the final conclusions of each part of the framework (WoE conclusions for E, A and T), add an additional level of clarity to the current WoE process. The latter may become more important as more and more compounds within a specific MoA (e.g., acetylcholinesterase inhibitors, etc.) are reviewed.

While only five chemicals are reported, the current framework does not appear to address how the WoE process will use OSRI data that falls outside of results for the E, A and T assessments. For example, in Chemical J, OSRI data were presented on delays in developmental effects in a mysid test. Where do these data fit within the framework in the absence of a Tier 1 screen for invertebrate hormonal pathways? How this type of data will be used in the final decision for a Tier 1 assessment is unclear and may become important, especially if more data are used from ASTER or other existing ecotoxicity EPA databases.

Charge 1.2. For each of the case studies, please comment on whether the performance criteria are clearly stated for the Tier 1 assays and, when results were not within the boundaries of the performance criteria, whether EPA has clearly expressed why the data are still considered reliable.

Panel Response:

In general, the discussion of performance criteria in the White Paper varied across assays and case studies, with more detail on specifics typically given for the *in vitro* assays. The specifics of the criteria for each assay were often not stated. It was assumed that when criteria were not specifically defined, they were met, although in some cases it was explicitly stated that all criteria were met. The description of the rationale for accepting data that fell outside the performance criteria also varied, although often it was clear that the discrepancies were minor and occurred in only one or a few endpoints so that they did not affect interpretation of the assay as a whole. Statements such as "performance criteria were generally met" or "performance was reasonable" often appear in the White Paper, but are not particularly descriptive. A consistent format for the statements of the criteria for each assay, whether the assays met the criteria, and why this was or was not a concern would be useful.

As discussed at the May 2013 SAP, the criteria provided with the guidelines were considered to be more general guidance than absolute performance criteria that needed to be met for assay acceptance. In the current White Paper, the assay criteria are sometimes described as

"requirements", "permissible", or "allowable" although they are, in fact, somewhat flexible and appear to be dependent on interpretation of the assay results as a whole. There was also discussion at the May SAP of reevaluating and revising the performance criteria based on the results from the entire initial set of compounds that went through the battery (e.g., setting acceptable CV% for organ weights based on historical lab data). This was again mentioned in the public comments at this meeting. The Panel encouraged the Agency to consider those recommendations and consider revising the criteria based on the screening results obtained thus far and more clearly establish what is in fact acceptable.

The comments below deal generally with the criteria stated for each of the Tier 1 assays for each of the Chemicals A, S, J, and N. Since the data presented for Chemical X were all based on OSRI rather than Tier 1 assays, performance criteria for the assays discussed was not presented. The main point of the listings for each chemical is to indicate what the Panel thought were inconsistencies across compounds in the manner in which the criteria were presented and in which the rationale for the acceptance of data that did not meet the criteria was described.

Chemical A

The issue of precipitation at the highest dose levels was discussed for several of the *in vitro* assays. In the cases of the ER binding assay and the aromatase assay, it was mentioned that testing of half-log doses, as recommended in the guidelines when precipitation is an issue, was not done and this contributed to considering the results of those assays "equivocal". Other deviations from the guideline criteria are described as minor and it was stated that the criteria were "generally met".

For the uterotrophic assay, the performance of the positive control was described as acceptable and it is also noted that the ratio of the uterine to body weight in controls was within guideline specifications. This is the only compound for which this was noted.

In the Hershberger assay, only one organ in one arm of the study was noted to be slightly out of range for acceptable CV%, and it is clear that this is considered minor.

For the pubertal assays, there was no mention of performance criteria. It was assumed that these were all met, but some statement should be included.

The description of the FSTRA assay indicates that several validity and performance criteria were not met (fecundity, spawning frequency, fertilization success, dissolved O_2 for a portion of the assay); however, it was stated that "performance was reasonable". This is an example of where perhaps the performance criteria need to be revisited since data outside the performance criteria are actually acceptable.

For the AMA, except for the indication that dose selection criteria were met, there was no discussion of other performance criteria.

Chemical S

For the *in vitro* assays, it was stated that the performance criteria were met with minor exceptions for estrogen receptor transcriptional activation (ERTA), steroidogenesis, and aromatase assays. ERTA and steroidogenesis deviations were considered minor. For the aromatase assay, it was indicated that 50% inhibition was not achieved, but that the results were considered acceptable because there was a dose response.

For the uterotrophic assay, the acceptable performance of the positive control was the only criteria discussed.

For the Hershberger assay, it was indicated that the CV% targets were met for all tissues for the vehicle/negative controls and high dose.

For the pubertal assays, there was no explicit discussion of criteria.

For the FSTRA assay, it was indicated that the % fertility was slightly below the criteria (93 and 90% versus 95%) for the negative and solvent controls. Dissolved O_2 dropped below criteria levels for an undefined period of time, but corrective action was taken, and there was inadequate reporting of secondary sex characteristics and clinical observations. The dose confirmation results had a CV% of >20%, but it was stated that there was no impact on study interpretation.

For AMA, exposure issues were noted and exposure described as inadequate. No other performance criteria were explicitly discussed.

Chemical J

For the *in vitro* studies, it was stated either that performance criteria were met (ER, ERTA) or that deviations were minor (AR, steroidogenesis, aromatase). For the AR assay, B_{max} and K_d were below the expected range, but there was generally good reproducibility that lessened the concern. Non-specific binding was outside of guidelines (24% versus 20%), but all other criteria were met. Steroidogenesis criteria were generally met, with some deviations in the levels of estradiol and testosterone produced and the level of inhibition of estradiol production by prochloraz. Minor deviations were described for aromatase in one of three assays.

For the uterotrophic assay, only the acceptable response of the positive control was discussed. No other criteria were mentioned.

For the Hershberger assay, it was stated that all criteria were met in terms of % coefficient of variation (% CV) for organ weights.

For the pubertal male assay, criteria were not discussed except that it was noted that the high dose was adequate. For the pubertal female, it was noted that the age and weight of vaginal opening were within performance criteria, but no other criteria were discussed.

For the FSTRA assay, it was noted that all performance criteria were met.

For the AMA, it was noted that validity requirements were met, but that there were not at least two concentrations that did not show overt toxicity.

Chemical N

For ER binding and ERTA, it was noted that all performance criteria were met.

For the AR binding assay, B_{max} and K_d were indicated to be lower than those in the validation assays and non-specific binding was 24.6% (guideline 20%) in one assay.

In the steroidogenesis assay, E2 levels were less than the recommended level of 40 pg/ml, but they were stated to be sufficiently above the detection limit to be useful.

For the aromatase assay, criteria were stated to be met except the top of the curve was 112% in one run (guideline 110%) and logIC50 for the positive control (4-OH-ASDN) was slightly outside the recommended range.

For the uterotrophic assay, the positive control response was stated to be adequate, but no other criteria were mentioned.

In the Hershberger assay, several organ weights exceeded the guideline CV%.

For the pubertal male assay, there was no discussion of criteria. For the pubertal female, all organ weights were stated to be within guideline CV% except for a minor deviation for the liver (13.29% vs. 13.13%).

For the FSTRA assay, it was stated that all validity criteria were met.

For the AMA, there was no direct discussion of criteria. It was indicated that there were some late stage individuals that were examined and an incidence of bent tail that was not considered to interfere with assay interpretation. There was discussion of nutritional and water quality variables.

Charge 1.3. The test guidelines for Tier 1 assays recommend that the organism is challenged by attaining sufficiently high treatment doses/concentrations. Difficult to test substances may be encountered in Tier 1 screening. Chemical S is an illustration of this situation. In the case of Chemical S, consistent exposure was not achieved in the Amphibian Metamorphosis assay (AMA) due to the physical-chemical characteristics of the test substance. The compound has low solubility and is highly lipophilic (high Kow) and prone to sorbing to surfaces (high Koc). Due largely to these properties, the contributing laboratory performing the AMA with Chemical S did not achieve a concentration level high enough to produce a response indicative of a maximum tolerated dose. Nonetheless, the Agency concluded that the data were still useful in the WoE analysis. This determination was based on the Agency's understanding that while measured exposure concentrations were lower than the targeted

nominal concentrations, exposure was reasonably quantified and that it is not likely that the chemical would be any less problematic to test if the study were repeated. Further, while higher exposure concentrations could have been achieved in the AMA, the FSTRA indicates that these higher concentrations likely would have resulted in overt toxicity.

Please comment on the Agency's conclusion regarding the utility of the AMA data for Chemical S to still reliably evaluate its potential endocrine interaction in a WoE analysis.

Panel Response:

The concentrations of Chemical S achieved in the aquarium water in the AMA were below the targeted nominal concentrations (as low as 8.8% of nominal for the highest test level). There was also large variability in the measured concentrations; the % CV ranged from 33-69% (the performance criterion maximum CV was 20%). Although not indicated in the White Paper, Chemical S was dissolved in dimethylformamide. This is the same solvent used for the Fish Short-Term Reproduction Assay (FSTRA), although in the FSTRA the recoveries were higher and the % CVs were lower. It is unknown why the investigators conducting the FSTRA were more successful in working with the chemical than those conducting the AMA.

There were no significant effects of Chemical S on growth and development or thyroid histopathology at the concentrations achieved in the AMA. Effects were reported with reference to the clean water (negative) control, not the solvent control. Solvent effects were reported for thyroid histopathology in the AMA and that complicates interpretation. The study author did not report clinical observations.

The Panel stated that it was difficult to directly answer this charge because statements in the White Paper and the preamble to the charge appear to be inconsistent. This made it hard to know what the Agency's conclusions are on this point. For example, in the White Paper the Agency concluded "The very low recoveries and reported instability (likely due to low solubility and high Koc) of Chemical S in the test system make the results of this amphibian metamorphosis study difficult to interpret." "Adequate exposure was not achieved in the Amphibian Metamorphosis assay (AMA), and therefore was not successfully tested in this assay. Testing in the AMA assay is limited by the physico-chemical characteristics of the test substance." "Overall, exposure levels in the AMA were determined to be inadequate, although there were no HPT-related effects within the range of concentrations tested." Thus, most discussion concerning results for Chemical S in the AMA indicates that the findings are difficult to interpret. However, in the preamble to the charge question the Agency states: "Nonetheless, the Agency concluded that the data were still useful in the WoE analysis. This determination was based on the Agency's understanding that while measured exposure concentrations were lower than the targeted nominal concentrations, exposure was reasonably quantified and that it is not likely that the chemical would be any less problematic to test if the study were repeated." For Chemical S, the discussion of the conclusion in section E of the White Paper that the data are still useful was not clear. It only appears in this form in the preamble to the charge question.

The Panel identified several problems with these conclusions. First, the Panel disagreed that the exposure was reasonably quantified, and that because of the low solubility and high

variability in the measured concentrations, the results do not "enable reviewers to determine responses" as stated in the White Paper. Second, even if one were to accept the very low recovery and high variability, and attempt to evaluate responses, it was not clear how the measured concentrations of Chemical S compared to expected environmental exposures. To evaluate the potential for endocrine disruption or toxicity, that is, whether the concentrations of Chemical S achieved in the AMA were relevant, it is important to know the range of exposure levels seen in the environment. As stated, Chemical S has low solubility. Therefore, it is likely that its concentration in aquatic systems would be very low, making it unlikely that exposure concentrations achieved in the AMA were ecologically relevant given the chemical's low solubility and propensity to partition to organic matter.

The Panel noted that the case study for Chemical S raised a larger question regarding the goals of the Tier 1 assays, in particular the two aquatic vertebrate assays (FSTRA and AMA). For example, the AMA is intended to serve two purposes: 1) to evaluate the potential for endocrine disruption in aquatic vertebrates (i.e., it is considered an ecotoxicological assay) and 2) using *Xenopus* as a model vertebrate to test for thyroid disruption. Both goals may be satisfied if the chemical is soluble in water. Neither goal will be satisfied if the chemical is insoluble as for Chemical S. This can lead to the loss of an important bioassay (perhaps two if the FSTRA is lost) from the Tier 1 battery to evaluate endocrine disruption, and the failure to address the ecotoxicological potential of the chemical for aquatic vertebrates (fish and amphibians). Chemicals with low aqueous solubility partition into organic matter and may bioconcentrate as aquatic organisms are exposed through food, not water. Such effects would not be captured in the FSTRA and the AMA.

The Agency may consider whether an alternate route of administration is possible for such chemicals. For example, the chemical may be injected or delivered in the diet. If this is not feasible, then perhaps the Agency should conclude, based on solubility data, that a chemical cannot be adequately tested in the AMA or the FSTRA via an aquatic exposure route. It may not be generally advisable to attempt to test chemicals with low aqueous solubility in these assays using the water medium route of administration. Determination of chemical solubility and possible routes for administration should be made in consultation with EPA scientists who are expert in determining chemical solubilities in aquatic systems (e.g., scientists at the Mid-Continent EPA Laboratory in Duluth, MN). These decisions should be made with knowledge of the likely route of exposure for aquatic species for a specific chemical.

Although it is difficult to interpret findings with Chemical S due to its low aqueous solubility, the Panel agreed with the Agency's conclusion that the assay should not be repeated, at least not using the same route of administration.

Lastly, the Panel disagreed with the Agency's statement, "Based upon WoE evaluation of EDSP Tier 1 data and OSRI, Chemical S does not demonstrate a potential to interact with the thyroid pathway." While the negative responses in the rat pubertal assays indicate that Chemical S does not interact with the thyroid hormonal pathway, the "inadequate" study findings for the AMA means the AMA response may not be considered "complementary and redundant" to the response of the rat pubertal assays. Also, the Tier 1 battery can only evaluate a limited number of thyroid endpoints, so any statements regarding thyroid disruption need to be qualified,

recognizing the limits to the assays which may not adequately address chemical effects on thyroid hormone metabolism or hormone action.

Charge 2. The WoE guidance decision framework is a hypothesis-based approach that promotes the analysis of effects at different levels of biological organization (molecular, cellular, tissue/organ, organism) using diverse in vitro and in vivo studies to determine whether or not a chemical interacts with E, A, or T signaling pathways. Pivotal to this approach is the concept of adverse outcome pathways (AOP) or modes of action (MoA). All of the case studies illustrate different data situations when characterizing the level of biological organization and the conditions under which potential endocrine interactions are expressed (e.g., only in organisms with intact HPG or HPT axes, coincident with overt toxicity, etc.) and varying degrees of complementarity and redundancy in responses encountered in the WoE analysis.

Charge 2.1. Chemicals do not necessarily act by one adverse outcome pathway (AOP). The case studies provide an illustration of competing AOPs (endocrine versus non- endocrine pathways or mechanisms of pesticidal mode of action that lead to overt toxicity). Although it is important to ensure adequate dosing in studies and reach some level of toxicity, overt or pronounced toxicity was relatively common at high treatment doses/concentrations in the Tier 1 and OSRI studies in these case studies. When overt toxicity is observed in the Tier 1 in vivo assays, it is an important objective to integrate the understanding of overt toxicity in the context of the apical responses associated with potential chemical interactions with E, A, or T signaling pathways. These case study analyses include a characterization of the treatment-dependent nature and severity of the overt toxicity as well as the specificity of the potential endocrine-related responses coincident with the overt toxicity. These analyses inform the weight that is placed on certain Tier 1 responses in the presence of overt toxicity. The case studies provide illustrations of the different situations encountered.

Charge 2.1.a. Chemical A can result in cholinergic toxicity given that its pesticidal mode of action is cholinesterase inhibition. In particular, overt toxicity was observed at high concentrations in the FSTRA. Although a number of endocrine responses were observed (e.g., decrease in female VTG, fecundity/fertility, GSI, male tubercles) at the highest concentration in the FSTRA, there was also pronounced overt toxicity that included abnormal behavior and significant body weight reductions consistent with cholinergic intoxication. Given the directionality of the FSTRA responses (i.e., decreases in the measured endpoints), EPA concluded that the effects found at the high concentration in the FSTRA may not necessarily be reflective of an endocrine-mediated response, but rather a reflection of a compromised organism with limited ability to maintain reproductive function and homeostasis. Although in male fish, overt toxicity was not observed at the intermediate concentration, possible endocrine responses were limited to two effects that lacked diagnostic specificity (i.e., altered GSI and histology).

Please comment on how the Agency has applied its decision logic to integrate an understanding of overt toxicity in the context of observed Tier 1 in vivo responses , and in particular, the Agency's determination not to place weight on the FSTRA high concentration responses coincident with overt toxicity.

Panel Response:

There are two places in the 2013 EDSP WoE White Paper that discuss decision logic for the FSTRA: Table 3 and Figure 3. Table 3, *Decision Logic for Interpretation of a Chemical's Potential for Endocrine Interaction for the FSTRA Assay*, is a general listing of FSTRA endpoints. Table 3 indicates that a "positive response" would be assumed for endpoints reported as statistically different compared to controls. It does not provide the diagnostic utility of the endpoints nor does it discuss how endocrine effects should be evaluated when overt toxicity occurs. Information from Table 3 applicable to the FSTRA is reproduced below:

Fish Reproduction (FSTRA)	Positive: Reported change in one of these endpoints (statistically significantly different from control ($p<0.05$)): Fecundity Fertilization success Secondary sex characteristics Gonado-somatic index (GSI) Histopathology of endocrine-dependent reproductive tissues (gonads) Plasma concentrations of vitellogenin (VTG) Plasma concentrations of sex steroids (optional)

Figure 3, *Decision Logic for the Interpretation of the FSTRA Results*, is part of the 2013 EDSP WoE White Paper for conducting the FSTRA test. The Agency should be commended for including this figure in the 2013 EDSP WoE White Paper because it provides very useful information on the diagnostic utility of the different endpoints and ranks them from greatest to least significance. The Panel recommended that the Agency provide similar decision logic figures for the other Tier I assays, especially when more than one endpoint is evaluated within an assay.

There are two main parts to Figure 3. The first part is determining whether adequate dosing was demonstrated (steps labeled Toxicity Assessment (Tier 1)). This takes into account dosing when overt toxicity occurs. The second part discusses the diagnostic utility of the endpoint. The Agency has successfully applied the correct decision logic for Chemical A based on dose selection and diagnostic utility for assessment of endocrine activity except for female dosing, as discussed below. The Panel recommended a more detailed discussion be included regarding this endpoint in the weight-of-evidence discussion for Chemical A. The conclusions from the decision logic for males and females are different.

Toxicity Assessment (Tier 1)

Females: At the high dose, there were non-lethal signs of overt toxicity: overt morbidity (abnormal behavior and swimming patterns), significantly decreased body weights and large

decreases in fecundity and fertility. At the medium dose, body weight was significantly decreased in females by 17% and that could be considered overt toxicity. It is unclear whether the Agency considered body weight decreases in females at the middle dose to be overt toxicity. The Panel recommended the Agency reevaluate the criteria used to determine whether overt toxicity occurred for specific endpoints (e.g., what level of body weight decrease in the FSTRA is considered overt toxicity).

There were no adverse effects at the low dose. When both the high dose and middle dose show signs of overt toxicity, Figure 3 indicates adequate dosing was demonstrated, but indicates to proceed with caution because endocrine effects are difficult to distinguish from toxicity.

Males: At the high dose, survival was not decreased, but there were non-lethal signs of overt toxicity: overt morbidity (abnormal behavior and swimming patterns). At the low dose and medium dose in males, there were no signs of decreased survival or signs of non-lethal overt toxicity. Based on the decision logic in Figure 3 this chemical would be evaluated for endocrine activity.

Evaluation of Endocrine Activity

Females: The only concentration where overt toxicity did not occur was the low dose. The only endocrine effect observed in females was a decrease in VTG by 62.5%, a very strong effect. The Panel recommended the Agency revisit the significance of decreases in VTG in females observed at the low dose, since this potentially could be a significant finding. According to Figure 3, decrease in VTG in females would be interpreted as "likely endocrine action".

Males: At the medium dose, GSI and histopathology were affected, whereas at the low dose, only histopathology was affected (i.e., spermatogonia). Both of these effects indicate a possible endocrine action, but are not pivotal effects and lack diagnostic significance. These are nonspecific effects and could be due to endocrine or non-endocrine mechanism.

Interpretation of Endocrine Effects when Overt Toxicity Occurs

At the high dose, results of the FSTRA test showed a decrease in female VTG, fecundity/fertility, GSI, and male tubercles when overt toxicity occurred. Should these effects be considered endocrine effects? Table 3 and Figure 3 do not provide decision logic to evaluate endocrine effects when lethality or overt toxicity occurs, but this topic was discussed in the White Paper. The 2011 WoE guidance document and the 2013 EDSP WoE White Paper states:

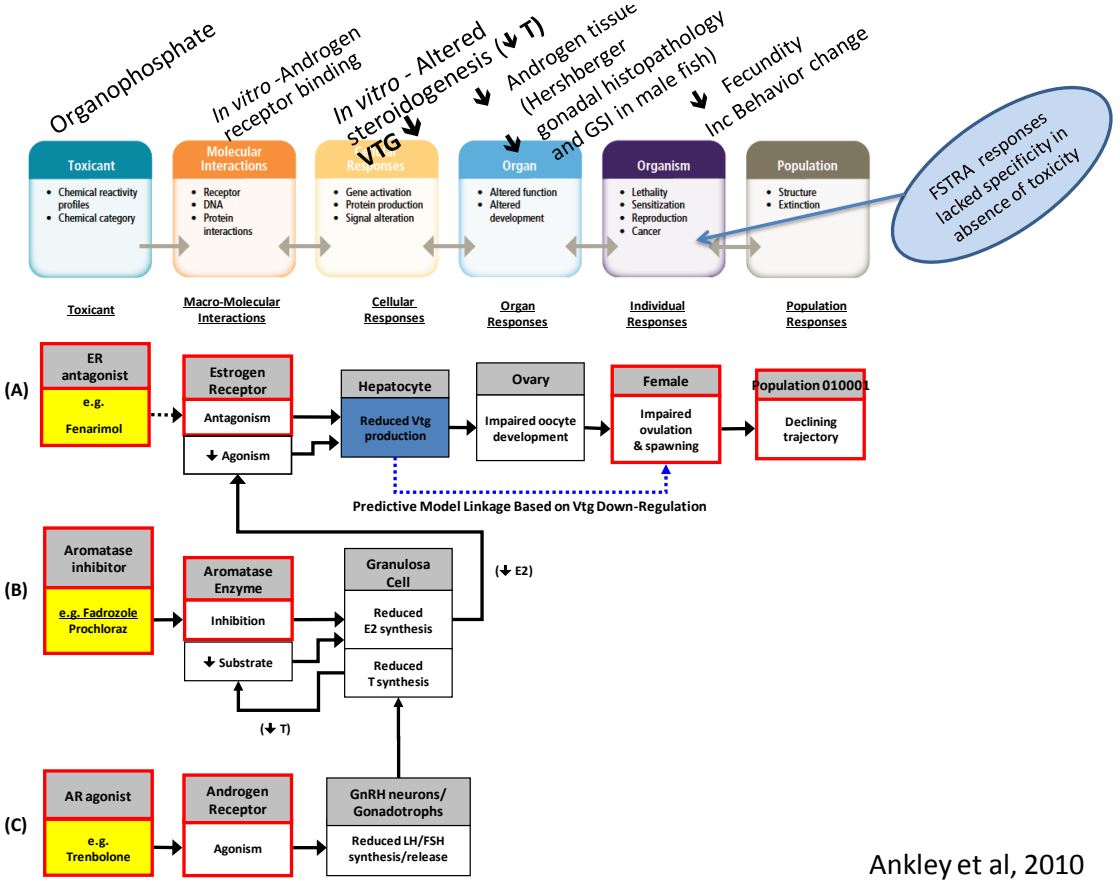
> *As indicated in the EDSP WoE guidance, "[i]t is important to consider and rule out other explanations for the observed results (e.g., secondary consequences of non-endocrine MoA or general toxicity) to the extent possible given the available data." It is realized that findings in the presence of overt toxicity may result in uncertainty as to whether or not the responses are mediated through an endocrine pathway. The current case studies provide examples of how endocrine responses that co-occur with systemic (overt) toxicity*

are evaluated and addressed on a case-by-case basis. These analyses consider the results from all of the Tier 1 assays and OSRI, as well as, the nature, directionality, magnitude, and dose trends of the Tier 1 endpoint responses. Additional consideration is given to the nature of the overt toxicity (e.g., mortality, reduction in body weight, behavior) and the treatment concentrations/doses at which toxicity occurred.

Based on the WoE guidance, the Agency has concluded that at the highest concentration, the observed endocrine effects were likely due to a non-endocrine MoA (i.e., cholinergic intoxication) and general toxicity and has discounted the endocrine effects that occur when overt toxicity occurs. The Panel agreed with this finding. In addition, the Agency also noted the directionality of the FSTRA responses supported this decision (i.e., decreases in the measured endpoints reflected a compromised organism with limited ability to maintain reproductive function and homeostasis). The Panel also agreed with this finding.

Generally, it is common practice in hazard assessment to discount effects that occur in the presence of overt toxicity as long as consideration is given to the nature of the overt toxicity (e.g., mortality, reduction in body weight, behavior) and the treatment concentrations/doses at which toxicity occurred. For example, in reproductive/developmental studies, treated pregnant dams frequently have greater than 10% body weight decreases relative to controls. If definitive teratogenic effects were observed in the fetus at doses that caused a 12-15 % decrease in body weight in the dams, developmental effects could not be discounted, but would require further evaluation.

Figure A. Different adverse outcome pathways affecting reproduction in fathead minnow.



Ankley et al, 2010

Figure A illustrates how mapping the chemical's known mode of action onto more detailed AOPs will make WoE analysis clearer. The top pathway (WoE/AOP) was presented as the summary of effects and impacts for Chemical A, a cholinesterase inhibitor. Pathways (A) through (C) represent different reproductive toxicity pathways in fathead minnow. Note that no comparison between known, or suspected, AOP for Chemical A is presented making determination of when overt toxicity occurs difficult.

Figure B. Mapping of WoE for Chemical A onto reproductive toxicity AOPs for fathead minnow.

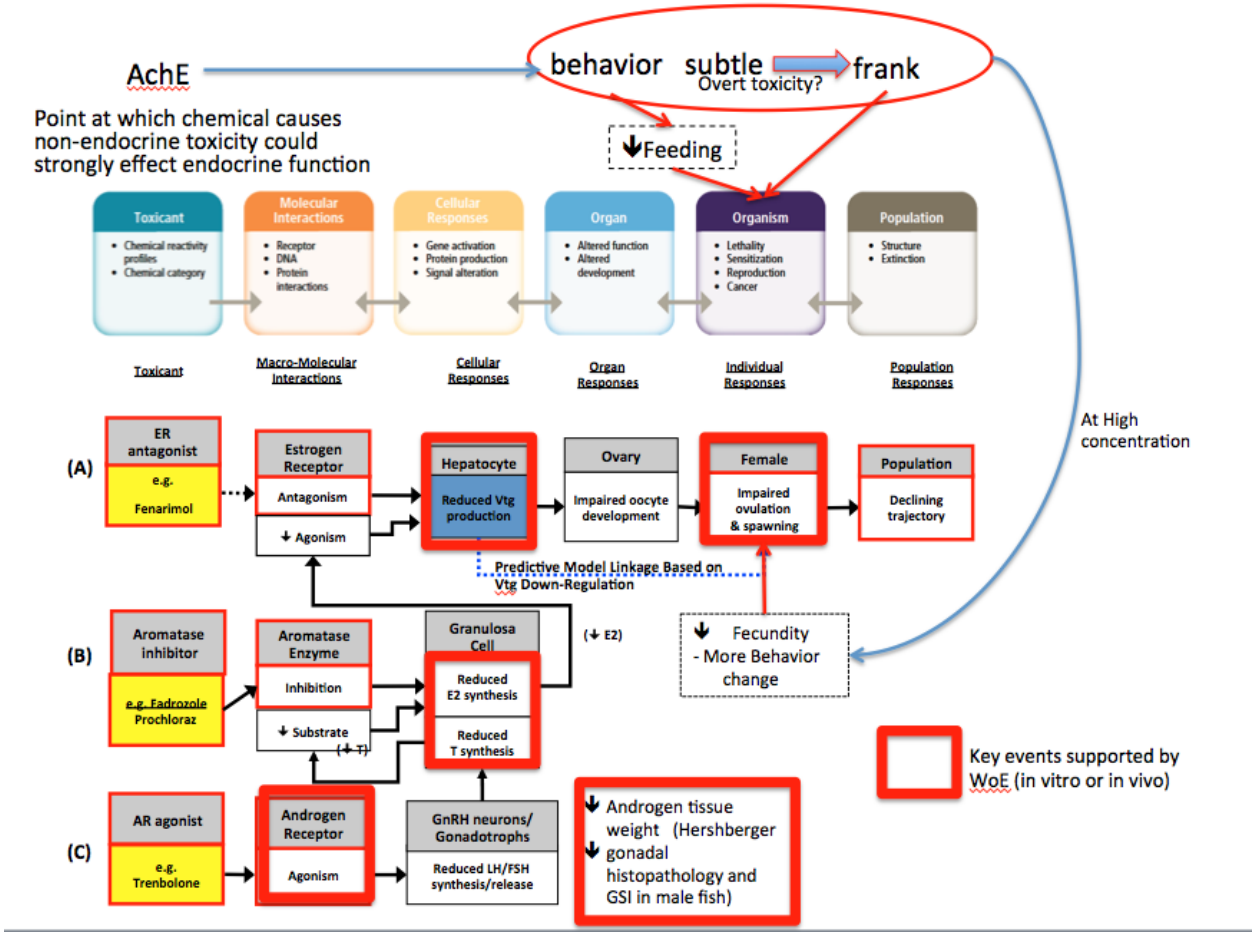


Figure B illustrates how mapping of WoE results onto AOPs for E, A, T and the non-endocrine mode of action for the chemical tested would help clarify WoE conclusions. For example, one could link behavioral changes due to cholinesterase inhibition to potential effects on feeding which could lead to reduced body weight, thereby affecting reproductive function.

One Panelist stated that the OSRI data provided clear evidence of adverse effects caused by acetylcholinesterase (AChE) inhibition at much lower doses (0.0032 mg/L) than were tested in the Tier 1 screening (0.011-0.82 mg/L). The White Paper reported adverse effects from the OSRI ecotoxicity data which indicated that, at doses as low as 0.0032 mg/L post-hatch, fat head

minnows exposed to Chemical A had scoliosis and reduced hatch of their progeny. Scoliosis will prevent fish from effectively feeding, which would, in turn, affect growth, development and reproductive output. Evidence of scoliosis at these low doses would constitute overt toxicity, which means that all of the effects observed in the FSTRA were potentially affected by overt toxicity at all doses. EPA was more conservative in their analysis and only excluded effects at the high dose as being affected by overt toxicity. EPA further concluded, that the remaining effects not impacted by overt toxicity at the intermediate and low doses (i.e., decreased VTG in females, decreased GSI and spermatogenesis in males) were suggestive of a cholinergic-compromised organism with a reduced ability to maintain reproductive function and homeostasis. The OSRI data mentioned above clearly support this conclusion by EPA. Further evidence of these effects is found in research on other organophosporus (OP) AChE -inhibiting pesticides, such as azinphos methyl, where fast swimming fish (e.g., red drum) had much greater sensitivity than slower swimming fish (e.g., mummichogs) as there were distinct differences in brain AChE and muscle AChE effects, with muscle AChE being more positively correlated with acute toxicity (Van Dolah, et al., 1997). These results indicate that there are distinct species differences in sensitivity to ACHE inhibiting chemicals; thus some species may exhibit overt toxicity while the same concentration in other less sensitive species may show no adverse effect. Thus, the metabolic activity level of the fish may play an important role in the fish response to chemicals with this MOA. Others have reported that significant inhibition of brain AChE in mummichogs was not correlated with alterations in metabolic rate, which would not lead to changes in growth and development (Fulton, 1989 and Fulton, et al., 1991). Thus, it is important to understand differences in tissue sensitivity in affecting AChE inhibition effects in accurately predicting subsequent effects on overt toxicity. If baseline metabolic rates (e.g., homeostasis as measured by oxygen consumption rate) were not adversely affected by OP exposure, it may be difficult to predict overarching metabolic effects within this AOP class of pollutants, given these species specific differences in interpretation of other FSTRA for other OPs.

Charge 2.1.b. The pesticidal mode of action of Chemical S involves the uncoupling of mitochondrial oxidative phosphorylation and resulting in the depletion of ATP. Another plausible mode of toxic action is related to its irritation properties including irritation that compromises the integrity of the gastro-intestinal tract in mammals leading to restricted caloric intake due to reduced food consumption. Reflective of these toxic modes of action, observations in the Tier 1 studies and OSRI included body weight reductions, behavioral effects, and decreased survival. The majority of potential androgen and estrogen-related responses (decreases in testosterone, decreases in male and female gonadal weights, delays in VO and PPS, decreases in male fertility, an increase in male GSI and VTG) were coincident with this overt toxicity. At concentrations where no apparent overt toxicity occurred, there were no endocrine related responses in the FSTRA, and responses in female rats were limited to a 2 day delay in VO, and for male rats, a decrease in the weights of two androgen-dependent tissues. The majority of Tier 1 responses were decreases in the measured endpoints, which were largely expressed in the presence of overt toxicity, are consistent with a depletion of ATP and restricted caloric intake. Although male VTG was increased in fish this is likely an artifact of a single elevated response.

Please comment on how the Agency's has applied its decision logic to integrate an understanding of overt toxicity in the context of observed Tier 1 in vivo responses, and in

particular, on the Agency's determination to place less weight on the Tier 1 in vivo responses in the presence of overt toxicity.
Panel Response:

Similar to 2.1a, the overall Agency interpretation of overt toxicity data as inadequate information for endocrine disruption was accepted by the Panel. There are clear issues with the overlap of alternate pathways to adequately evaluate endocrine disruption. The current methodology of Tier 1 testing does not allow for adequate interpretation of endocrine disruption at these overtly toxic levels. It should be noted that the results of *in vivo* studies at these overtly toxic levels are of clear importance to demonstrate that concentrations are being tested at a sufficient level to demonstrate an adequate effect of the chemical in question.

That being stated, this brings into question the logarithmic scales in which the Tier 1 testing is being performed in some assays, such as the FSTRA and the AMA. Tests such as the Hershberger with closer dosage intervals might be of less concern, but, nonetheless, decisions might be based on a single concentration. This is *de facto* what would occur in pubertal tests if the high dose demonstrates overt toxicity.

Ideally, concentrations just below the overt toxicity concentrations can be included without immediately expanding to Tier 2 tests to provide suitable information on screening possible endocrine effects of the E, A, or T pathways.

In lieu of the availability of these data, any effects that are observed at the concentrations below overt toxicity should be given adequate consideration. That is, if a single endpoint at a concentration below overt toxicity provides evidence that is consistent with endocrine disruption, this can merit special consideration for further testing at the Tier 2 level. In addition, consideration of both Part 158 toxicity data and the OSRI literature can provide additional important information on potential endocrine disruption that might occur at these intermediate levels.

As discussed in the previous charge question, the FSTRA flow chart decision tree is a concrete way to evaluate the data. While the remaining *in vivo* tests have fewer endpoints, a similar decision tree would be helpful to evaluate the potential for interactive effects of the E, A, and T pathways. A part of this decision tree might include the definition or potential characterizations of overt toxicity in these studies as they can strongly influence the interpretation of data. The measure of it in different organisms might be problematic. This was mentioned in the response to Charge 1.1.

Addressing Chemical S in particular, there are clear issues of solubility that are summarily addressed in the response to Charge 1.3. That being stated, each of the resulting effects found at the high concentration are consistent with the proposed toxic MoA (decreases in testosterone, decreases in male and female gonadal weights, delays in VO and PPS, decreases in male fertility, and increases in male GSI and VTG). The Panel thought that these results were potentially the result of a toxic level of exposure and, therefore, did not provide sufficient evidence for endocrine disruption on their own.

Other findings at levels that do not show overt toxicity, in the case of Chemical S, do show the potential for interaction with the androgen pathway. This is demonstrated via the decrease in weights of two androgen-dependent tissues at the lowest concentration in the Hershberger assay. The Panel agreed with this assessment. In the case of potential estrogen pathways, there was a positive endpoint (2-day delay in vaginal opening) in the pubertal assay at the low concentration. The Agency declared that the WoE does not support a potential interaction of the estrogen pathway with this single indicator. However, the potential issue is that there might be affected endpoints in the FSTRA between the overtly toxic levels and the next lowest dose tested, which was a full 10-fold below the overtly toxic dose. The low solubility of Chemical S causes other limitations on the absolute number of endpoints available and reduces the ability for repeatability to be demonstrated. Therefore, while removal of overtly toxic levels and limited solubility might be adequate, the Agency should strongly consider such unique situations with caution. The inclusion of more information and other concentrations, or additional tests might be particularly merited in similar situations. That being said, with regard to Chemical S, this seems a reasonable judgment given the results of the remaining battery of tests and information from the OSRI, and the Panel agreed with the overall assessment of Chemical S.

The situation for Chemical A, as discussed in the response to the previous charge, is not as clear. The Panel already indicated the difficulty in evaluating the qualitative assessments. In conclusion, the Panel accepted the Agency's determination to place less weight on Tier 1 *in vivo* responses in the presence of overt toxicity. The Panel did express concern that the ability of the Tier 1 tests to detect endocrine disruption might be compromised dramatically when levels that express toxic responses are removed from consideration. This is of particular concern in situations that leave only a single test level or where the space between concentrations is an order of magnitude in difference.

Charge 2.1.c. Chemical N is a cyclic unsaturated ketone whose acute mode of toxic action is nonpolar narcosis (toxicologically induced and reversible stages of neural disruption, i.e. general anesthesia). Unlike the other case study chemicals, there is no pesticidal mode of toxic action for N given that it is an inert ingredient. Testing required reaching limit doses/concentrations in order to sufficiently challenge the animal. Potential androgen responses only occurred in the FSTRA (decrease female VTG, decrease fecundity/fertility, altered histology) and in the male pubertal assay (decreases in testosterone, decreases in androgen sensitive tissue weights, delays in PPS) near limit doses/concentrations (as described in the white paper and test guidelines). However, a significant decrease in female VTG was observed at the intermediate dose. Observations of overt toxicity (decreased body weights and feeding) were reported in the highest treatment group (i.e., near limit concentrations) in the FSTRA, but no overt toxicity was reported in the male pubertal assay. Unlike Chemicals A and S, the overt toxicity is not as pronounced for Chemical N. The responses in fish and rats at the high dose could be due to a compromised metabolic ability and inability to reduce chemical load.

Please comment on the Agency's analysis in characterizing Tier 1 responses that are expressed at or near limit doses where some degree of overt toxicity occurs, and the extent to which such responses are considered in the WoE analysis.

Panel Response:

As stated previously, the Agency put less weight on results from assays where overt toxicity occurred, and the Panel agreed with this approach. Tier 1 EDSP results for any test substance dose that causes overt toxicity will pose significant interpretational problems, since this disrupts physiological and endocrine functions in a manner that alters many of the parameters measured in the assays. In addition, use of one or more overtly toxic doses in the EDSP assays, which are conducted with a small number of dose groups that are generally widely spaced, can result in the failure to detect a true high dose endocrine response. Moreover, such doses can potentially comprise a significant homeostatic challenge that triggers the activation of hypothalamic-pituitary-adrenal axis (i.e., stress axis) responses that can also impact parameters being measured. In general, the Panel thought that the Agency should avoid testing at or near limit doses, or at doses shown to be overtly toxic in range finding studies, in order to better detect true endocrine-mediated high dose toxicity. This could be done by more testing at lower doses; in particular to the FSTRA, it was noted by a Panel member that due to the logarithmic dosing paradigm, the Agency has ample room to try a range of lower doses without adding a significant amount of time to run the assay.

Chemical N is a good example of an agent that had effects at toxic doses, as anti-androgenic and estrogenic actions occurred in the FSTRA, but with one exception, this was only at overtly toxic doses. Chemical N also caused anti-androgenic effects in the male pubertal assay at high doses, but although the dose was near the maximum tolerated, signs of overt toxicity, reportedly, did not accompany these findings. Since there appears to be no effect of Chemical N on steroidogenesis or androgen receptor binding, as noted by the Agency, a possible cause of the anti-androgenic effects in the male pubertal assay may be an action on upstream HPG parameters that indirectly drive testosterone synthesis. As noted by the Panel, the effects of Chemical N could also be due to stress, or some other non-endocrine effect. As such, and as recommended in the May SAP, it would be important in such instances to take into account measures of hypothalamic-pituitary-adrenal axis factors like corticosteroid concentration in the presence of Chemical N to clarify the role of stress in its actions. Moreover, since the chemical appears to work upstream of the gonads to disrupt HPG function, the Agency should also consider measures such as gonadotropin releasing hormone and/or luteinizing hormone release that would confirm or refute this.

In conclusion, the Panel acknowledged that the EPA is in a difficult position in interpreting the EDSP assays, as it is important to determine maximum tolerated doses for test chemicals, while at the same time avoiding the confounding effects of toxicity and stress on the parameters measured. Overall, the Panel concluded that assays in which there is overt toxicity are not useful for interpretation of whether a compound has an endocrine effect, and that results for assays using doses near the test limits need to be evaluated carefully.

Charge 2.1.d The case study analyses described above all involve situations in which overt toxicity was observed coincident with Tier 1 responses.

Please comment on the Agency's overall approach to characterizing Tier 1 responses coincident with overt toxicity and determining the weight to be given to such responses.

Panel Response:

The Panel and EPA are in agreement that overt toxicity is a confounding variable. Comments by Panel members and the public indicated that even more consideration should be taken in evaluating Tier 1 endocrine endpoints in relationship to non-endocrine toxicity and when selecting the highest dose/concentration to be used in Tier 1.

Evidence of endocrine toxicity in the presence of compound-induced toxicity should be highly scrutinized and suspect. When using a weight-of-evidence approach, all available toxicological data should be used to put findings in endocrine sensitive tissues/organs into context. Toxicity in other major organ systems can affect Tier 1 study endpoints, for example;

- Gonadal organ weights should be normalized when significant changes in body weight are noted; it is also important to analyze if decreased body weight gain occurs concurrently with decreased feed intake which may result from palatability issues rather than toxicity.
- Hepatic enzyme induction, hepatotoxicity, and nephrotoxicity can affect metabolism and clearance of compounds as well as endocrine hormonal levels.
- General morbidity and neurotoxicity can indirectly affect reproductive performance.
- Stress can adversely affect feedback regulated systems.

Dose range finding studies or some other dose/concentration titration paradigm should be used to determine the exposure/dose-limiting toxicity to avoid confounding interpretation of endocrine related effects at overtly toxic exposure levels. EPA should standardize the criteria used for determining overt toxicity that confounds interpretation of data (e.g., morbidity, mortality, decreased body weight gain with or without concurrent decreases in feed consumption, and exposure limiting non-endocrine organ toxicity such as neurotoxicity, cardiotoxicity, hepatotoxicity, nephrotoxicity, etc.). Dose-limiting toxicity should be based on the most sensitive endpoint and used to inform selection of final doses. EPA should consider using a minimally toxic dose (LOAEL) rather than a maximum tolerated dose (MTD) for the definitive assays/studies. ORSI data may be helpful in defining a dose-limiting toxicity.

In summary, the Panel agreed that little weight, if any, should be placed on signs of endocrine disruption in the presence of overt toxicity. All effects in endocrine sensitive tissues should be evaluated in terms of primary interactions with the endocrine system vs. secondary effects related to toxicity in non-endocrine organs or overall disruptions in homeostasis.

Charge 2.2. In certain case studies, there was a lack of anticipated complementary and redundant responses (within an in vivo assay or across assays) at different levels of biological organization (molecular, cellular, tissue/organ, and organism) indicative of a chemical interaction with an endocrine signaling pathway. The estrogen signaling pathway will be used as an illustration. In the case of Chemical N, the mammalian assays were negative and responses within the FSTRA did not progress to higher level responses (e.g., an effect on VTG did not translate to an effect on gonadal-tissue or on fecundity). In the case of Chemical A, the rat uterotrophic and female pubertal assays (i.e., an organism with an intact hypothalamic-pituitary-gonadal axis) were negative for estrogen-related responses, and although there were some responses in the FSTRA in the absence of overt toxicity, they lacked diagnostic specificity (e.g., effects on male gonadal tissue or GSI). Given the lack of complementarity and redundancy in responses within and across assays, the Agency considered these situations as insufficient to support a robust conclusion of an interaction with endocrine signaling pathways.

Please comment on the decision logic the Agency has used to characterize these types of situations where there is a lack of robustness in terms of complementarity and redundancy, and the transparency and reasonableness of the approach.

Panel Response:

The current WoE review process focuses on the completeness and consistency of the MoA/AOP pathway and helps define whether the pathway is complete from lower levels of biological activity (e.g., E, A or T receptors at a cellular level = *in vitro* assays) to higher levels of biological activity (e.g., tissue, organ, organ system, organism level = *in vivo* assays) across multiple taxa (mammals, fish and frogs). Included in this consistency of response is an assessment of the complementarity (e.g., consistency within an assay with multiple interrelated AOP endpoints for E, A and T) and redundancy (e.g., consistency among multiple lines of evidence from the different assays with common and interrelated AOP endpoints for E, A, and T). WoE is not a prescriptive process but an integrative and interpretive process routinely used by EPA to evaluate health and ecological information for other types of environmental assessments and its approach in the EDSP is appropriate. As such, it is currently a qualitative process for describing the degree of consistency within the Tier 1 screening data specific to each of the three vertebrate endocrine pathways.

In many respects, the WoE approach is consistent with conventional risk assessments in terms of defining exposure pathways, although the WoE for the EDSP is not a risk-based approach *per se*, but rather an integrative, consensus based approach for interpreting multiple lines of evidence within a screening assay battery. In a conventional risk assessment framework, if the exposure pathways are incomplete, there is no risk. Similarly, in an AOP exposure framework, if there is a lack of complementarity within a single species bioassay or a lack of redundancy for multiple species bioassays for the 3 endocrine pathways, there is an incomplete AOP and a perception of less or no endocrine disrupting chemical risk. The WoE approach for evaluating potentially endocrine disrupting chemicals uses a much different approach to derive its final integrative conclusions, drawing on best professional judgment using decision logic trees as part of the process. The decision logic tree for the FSTRA is an excellent example of this

approach and a similar decision tree approach presented for the AMA was useful. The Panel recommended using a decision tree approach for each of the three endocrine pathways (E, A and T) within the full battery (rodent, fish and frog) of *in vivo* assays. The WoE approach also uses an adopted subset of Bradford Hill Criteria including: (1) sequence of key events; (2) coherence; (3) strength and limitations; (4) consistency; and (5) biological plausibility in evaluations of complementarity and redundancy. Using this approach, no single study or endpoint is generally expected to support a regulatory decision by itself; rather, it is the totality of the full battery that becomes important in final decision-making. EPA's introductory presentation described the battery of assays for E/E-; A/A-, Steroid Synthesis of T/E; HPG and HPT with the number of assays for each of these eight different endocrine pathways ranging from a minimum of a battery of three bioassays to a battery of five bioassays. All of the assay batteries, except for the HPG and HPT, include at least one *in vitro* assay along with several *in vivo* assays. This design is critical as it allows for assessments of chemicals directly (e.g., *in vitro* and *in vivo* assays), as well as metabolites (*in vivo* assays), to make these important distinctions. This design also allows for assessments with increasing biological complexity (cell → tissue → organ → organism) to be made. The completeness of this response along a specific endocrine pathway within and among bioassays is what defines consistency within a battery as measured by complementarity and redundancy, respectively.

The White Paper described instances of a lack of consistency for a battery of assays, resulting in decisions that excluded some data as providing supporting evidence for endocrine disrupting effects within the battery. The lack of specific complementarity and redundancy in responses within and across assays often resulted in EPA concluding that these situations were insufficient to support a robust conclusion of an interaction with endocrine signaling pathways. For example, for the estrogen signaling pathways for Chemical N, the mammalian assays were negative and responses within the FSTRA did not progress to higher-level responses (e.g., an effect on VTG did not translate to an effect on gonadal tissue or on fecundity). Thus, there was a lack of evidence of specific estrogen binding in the *in vitro* assays along with a lack of complementarity in the FSTRA and a lack of redundancy between the FSTRA and mammalian assays in the *in vivo* assays. For the androgen signaling pathway, the *in vitro* assays and the Hershberger assay results show no AR-mediated androgenic or anti-androgenic effects, while the *in vivo* results showed some complementarity in potential anti-androgenic response in the mammalian bioassays (e.g., decreased testosterone, decreased seminal vesicle weight, and slight increase in age at PPS), but only at the highest dose. The FSTRA was generally absent of complementarity as only an effect on VTG at the intermediate dose was observed, with no redundancy with the mammalian assays. EPA also concluded that VTG could be affected by alterations in feeding behavior, which were not observed at lower doses. All assays for thyroid effects were negative for Chemical N, which was highly complementary and highly redundant for "no effects" in the *in vivo* assays. Thus, EPA's conclusions for Chemical N were: (1) estrogen pathway: no evidence of receptor-mediated interaction, possible downstream anti-estrogenic effects in the *in vivo* FSTRA; (2) androgen pathway: no evidence of receptor-mediated interaction, possible downstream effects in the mammalian male rat (anti-antigenic) and FSTRA (androgenic/anti-androgenic); and (3) thyroid pathway: no effects. Most Panel members generally agreed that these findings by EPA followed the decision logic within the White Paper and conclusions were supported by OSRI data and other outside references (e.g.,

feeding effects on VTG in FSTRA) that were peer reviewed and quality assured for inclusion in this process.

In the case of Chemical A, EPA concluded that, based on the WoE for Tier 1 and OSRI data, Chemical A does not demonstrate potential to interact with the estrogen pathway. Review of the Tier 1 data indicated that: (1) *in vitro* tests: ER binding and aromatase were equivocal while the ERTA and steroidogenesis assays were negative; and (2) *in vivo* tests: the uterotrophic and female pubertal assays (i.e., an organism with an intact hypothalamic-pituitary-gonadal axis) were negative for estrogen-related responses, and while there were some responses in the FSTRA in the absence of overt toxicity, they lacked diagnostic specificity (e.g., effects on male gonadal tissue or GSI). Results in the FSTRA were complicated by evidence of overt toxicity. In the FSTRA, the following effects were seen: (1) decreased GSI in males at the high and moderate dose; (2) altered histopathology (spermatogenesis) in males at the high, moderate and low doses and follicle effects in females only at the high dose; (3) fertility and fecundity were reduced only at the high dose; (4) changes in secondary sex characteristics (tubercle score) in males only at the high dose and (5) VTG was decreased in females at the high and low dose. Effects in the high (0.82 mg/L), intermediate (0.12 mg/L), and low (0.011 mg/L) concentrations in the Tier 1 FSTRA had direct evidence of overt toxicity (altered swimming behavior and body weight reductions) and in the OSRI ecotoxicity data, fathead minnows exposed to Chemical A at doses as low as 0.0032 mg/L post-hatch had scoliosis and reduced hatch of their progeny. Scoliosis will prevent fish from effectively feeding; therefore, it would suggest that reduced hatch was related to the overtly toxic effect of scoliosis that reduced the feeding efficiency of the fish that, in turn, affected resulting reproductive output. In questions to EPA during the presentation, there was agreement that evidence of scoliosis at these low doses would constitute overt toxicity, which results in all of the effects in the FSTRA being affected by overt toxicity at all doses. EPA was more conservative in their analysis and only included effects at the high dose as being affected by overt toxicity. EPA concluded that the remaining effects not affected by overt toxicity at the moderate and low doses (e.g., decreased VTG in females, decreased GSI in males and spermatogenesis in males) were suggestive of a cholinergic-compromised organism with a reduced ability to maintain reproductive function and homeostasis. The OSRI data referred to above clearly support EPA's conclusion to exclude these FSTRA effects as supporting an estrogen-related effect. This lack of complementarity and redundancy in the *in vivo* assays was affected by the confounding nature of overt toxicity for this chemical, which led EPA to conclude via a WoE approach that there was a lack of consistency within these parts of the Tier 1 battery. Effects of other AChE-inhibiting OPs in fish have been assessed by several researchers (Fulton, 1989; Fulton et al., 1991; and Van Dolah et al., 1997) which illustrate the importance of species differences, differences in levels of brain and muscle AChE, and lack of AChE inhibition causing effects on homeostasis as measure by general metabolic rate. This type of data summarized through OSRI will be important in future assessments of other AChE-inhibiting OPs.

This neurological effect caused by AChE inhibitors' MoA, thereby causing overt toxicity, raises the issue of separating neuro-endocrine coupling as part of the normal physiology of vertebrates. AChE inhibitors target the brain and other parts of the central nervous system which regulate and control endocrine functions of E, A and T. This issue was addressed in the EDSTAC Report and subsequent publications summarizing the workshop (Kavlock et al., 1996).

The neurological, immunological and carcinogenic work groups at EDSTAC all noted the complexity of identifying whether effects of xenobiotics on those systems were the result of primary or secondary aspects of endocrine disruption. Endocrine disrupting chemicals may adversely cause effects that can arise from either primary or secondary disturbances of endocrine function. A direct-acting endocrine disruptor affects the endocrine system first, which in turn results in toxicity in other organ systems. Conversely, an indirect-acting endocrine disruptor may affect a non-endocrine systemic target organ first, causing overt toxicity which may in turn influence the endocrine system to cause secondary neurotoxicity, reproductive toxicity, and/or immunotoxicity. It is very difficult, if not impossible, to separate the endocrine system effects from systemic target organs effects caused by overt toxicity, so this distinction should be viewed in an abstract manner. For these reasons, identifying agents as direct or indirect endocrine disruptors is problematic, and necessitates an integrated research approach such as the WoE approach. The EDSTAC Panel members concluded the complexity of this neuro-endocrine coupling would complicate results in future endocrine disrupting chemical assessments. This example with Chemical A clearly illustrates this complexity and it will be very obvious as other OP AChE inhibitors are assessed that the complete AOP battery results will provide more insight on the importance of this issue of overt toxicity affecting WoE assessments for endocrine disrupting chemical effects. It will be interesting to compare results for other classes of AChE inhibitors (e.g., carbamates) and for compounds such as fenoxycarb, a non-AChE-inhibiting insect growth regulator, as this may allow further understanding of the interactions between AChE AOP and endocrine disrupting chemical AOPs.

The dose design of the FSTRA (log spaced dose) and other parts of the Tier 1 battery may complicate results for compounds, such as Chemical A, which appear to have a high slope in the dose response curve for many end points. Examining OSRI data on the slope of the dose-response curve, by use of the ASTER database, for example, will be advisable as part of data review prior to starting Tier 1 battery tests.

Using this approach would seem to provide EPA the following possible conclusions that it may use in deciding on the degree of consistency within the battery of data: (1) positive *in vitro* MoA/AOP assay along with positive *in vivo* assay evidence of both complementarity and redundancy = Gold Standard; (2) positive *in vitro* MoA/AOP assay along with either positive *in vivo* assay evidence of complementarity or redundancy = Silver Standard, or (3) either (A) a positive *in vitro* MoA/AOP assay along with only a single positive *in vivo* assay endpoint with no evidence of complementarity or redundancy or (B) negative *in vitro* MoA/AOP assay along with positive endpoints *in vivo* assays that are non-specific to MoA/AOP with no evidence of complementarity or redundancy among these end points = Speculative Standard requiring additional testing to resolve the Tier 1 battery. Additional information on development of similar ranked or tiered approaches for the WoE approach was presented during the public comment opportunity by the Endocrine Policy Forum. The Panel recommended that EPA consider these suggestions in adopting or developing similar approaches to better define the WoE evaluation process and move toward one that is more quantitative.

Most Panel members thought that one of the greatest challenges is that, absent weighting of the power of individual assays, one is left with a counting exercise to document how many assays were mutually supportive for any given endocrine effect (E, A, or T). Some Panel

members proposed that certain *in vivo* tests are, in medical terminology, "pathognomonic" of certain hormone effects. Among these, a uterotropic response explicitly implies some excess estrogenic activity, and changes in male secondary sex organs in the Hershberger assay explicitly implies either pro- or anti-androgenic activity. Other endocrine outputs, especially for the thyroid hormone, do not benefit from equally precise *in vivo* responses.

Finding unambiguous effects in particular *in vivo* assays, in the absence of any overt toxicity, might be considered sufficiently powerful evidence so as not to rely on additional complementary or redundant evidence. In particular, these diagnostic *in vivo* responses explicitly do not rest on any *in vitro* or mechanistic evidence (receptor binding, transcriptional assays, etc.). If there is a clear uterotropic effect, no amount of negative evidence in the ER binding, steroidogenesis, or aromatase data is sufficient to negate this finding simply because of lack of coherence or complementarity. Given that there are reasons to prefer assays other than *in vivo* tests (costs, animal usage, etc.) it is important to reassert the diagnostic power of classic and well-understood *in vivo* tests.

Charge 2.3. In contrast to the situation described in question 2.2., Chemical J appears to interact with the estrogen signaling pathway in terms of complementarity and redundancy across multiple levels of biological organization as evidenced through altered steroidogenesis, resulting in decreased VTG in female fish which in turn translates to a higher-level response (e.g., reduced fecundity) in fish. However, this biological continuum was not observed in the Tier 1 rat female pubertal assays and the Part 158 mammalian data.

Please comment on the how the Agency has characterized this endocrine interaction at different levels of biological organization across taxa, and the transparency and reasonableness of the conclusions drawn. Please include in your response, comments regarding the Agency's conclusion about differences in sensitivities between taxa (i.e., fish and rats), regarding chemicals that appear to alter steroidogenesis.

Panel Response:

The Panel concurred with the Agency that Chemical J's interaction with the estrogen signaling pathway does not include interaction with the estrogen receptor (based on the ER binding assay in the rat uterine cytosol assay and the ERα transcriptional activation assay in the human cell line HeLa 9903). The Panel also concurred that two of the *in vitro* assays in the Tier 1 battery for Chemical J (i.e., the steroidogenesis assay (human cell line H295R) and the aromatase assay (human recombinant microsomes)) indicated interaction with the estrogen signaling pathway via effects on steroidogenesis. Specifically, in the steroidogenesis assay, at 100μM Chemical J treatment, testosterone levels were significantly ($p<0.05$) decreased, while estradiol levels were decreased ($p<0.05$) at concentrations of 0.1-100μM. In the aromatase assay, the 10^{-6} M concentration reduced aromatase activity by approximately 75%, and the 10^{-4} and 10^{-3} M concentration levels reduced aromatase activity by >99%.

Some of the *in vivo* tests in the battery also have endpoints linked to estrogen signaling; either direct measurement of estradiol (e.g., the optional measurement of steroid levels in the FSTRA) or indirect measures downstream of estrogen (e.g., the organ/structural changes in the

pubertal female assay in the rat and VTG levels in the FSTRA). In the pubertal female rat assay, Chemical J had no significant treatment-related changes or alterations on body weight gains, age of attainment of VO, body weight at VO, mean age at first estrus, mean cycle length, percent cycling, percent regular cycling, and organ weights. In the FSTRA, the highest treatment group (3.3 mg a.i./L of Chemical J) showed a significant ($p<0.05$) decrease in fecundity and female plasma VTG and estradiol levels, along with a significant increase in male and female GSIs. Alterations in male and female gonadal histopathology were also reported at 3.3 mg a.i./L.

As noted by the Agency for Chemical J, for effects via estrogen signaling (e.g., depressed estradiol production *in vitro* and decreased female VTG and estradiol levels *in vivo*), there does appear to be complementarity and redundancy of the *in vitro* and FSTRA tests in the battery. The *in vivo* data are of particular value in identifying the potential of a chemical to interact with estrogen, androgen, or thyroid signaling pathways as these data integrate the exposure, uptake, metabolism, and effects at various levels of biological organization, including at the whole organism level. Some endpoints in the *in vivo* tests are pivotal in expressing potential interaction with the endocrine pathways, for example VTG, and are therefore of particular value in understanding the potential for an adverse outcome. In the case of Chemical J as tested in the FSTRA, decreases in estradiol levels were complementary with depressed VTG, increased female and male GSIs, altered gonadal histology, and decreased fecundity. However, the redundancy of the results between the *in vitro* endpoints and the female fish in the FSTRA were observed neither in the rat female pubertal assays endpoints when tested at doses up to and including 400 mg/kg/day nor in the Part 158 mammalian data.

Given the lack of redundancy between the FSTRA and mammalian studies in the Tier 1 and OSRI for Chemical J, consideration of differences among taxa is warranted as relates to toxic effects of contaminants on the vertebrate endocrine system. This topic has been covered recently in the open literature as noted in the White Paper. Based on a review of the open literature, four known inhibitors of steroidogenesis (ketoconazole, fadrozole, fenarimol and prochloraz) were assessed by Ankley & Gray (2013) for activity across the FSTRA, uterotrophic, Hershberger, pubertal female and pubertal male rat tests. Three key issues are raised by the authors pertinent to Chemical J and why there can be differences in whole-organism estrogen signaling responses between the FSTRA and the mammalian tests, namely: (1) exposure route (water-borne vs. oral); (2) metabolism (e.g., first-pass hepatic metabolism when exposure is oral), and (3) greater specificity of the FSTRA than mammalian tests (including the female rat pubertal assay) in detecting inhibitors of steroidogenesis. Therefore, although there is a high degree of conservation of the HPG axes among vertebrates supporting the current WoE approach, differences in life history and taxa physiology as they relate to contaminant exposure and effects among taxa, do need to be considered when interpreting the Tier 1 data such as those obtained for Chemical J.

In conclusion, the Panel agreed that there is relevant and consistent evidence, if one focuses on the FSTRA, that Chemical J has the potential to interact with the normal function of the estrogen signaling pathway across different levels of biological organization, including cellular, tissue/organ and organism levels. This conclusion is made despite the potential for confounding effects across various levels of biological organization. It may be that the taxa differences (i.e., lack of effects in mammalian *in vivo* studies) are based on differences in taxa

sensitivities in regard to altered steroid synthesis (as discussed previously). The taxa differences are noteworthy and do underscore the need to consider factors such as route of exposure, uptake, metabolism, taxon physiology, etc., in making conclusions about the conditions under which specific chemicals perturb the endocrine system in organisms with intact HPG axes, including the role of overt toxicity in data interpretation. It also adds to the arguments that the totality of the data need to be considered, and that no one test in the battery is adequate in a WoE approach.

As an additional side note related to the analysis of Chemical J's effects, while the Agency concluded that a statistically significant ($p<0.05$) decrease of approximately 9% in female body weight at 3.3 mg a.i./L was not considered to be biologically significant, some caution should be exercised in the FSTRA results for females in this highest treatment group which is where the majority of the positive *in vivo* estrogen signaling data are found. Supporting the Agency's conclusion that the decrease in body weight is not biologically significant, the changes in body weight were not mimicked in the males and do not appear to have led to an overall decrease in general health status or performance of the test (as per the information provided in the White Paper). Panel members noted that if body weights per each treatment replicate (i.e., tank), or even for individual fish pre-exposure and post-exposure are not monitored in the FSTRA, it is a design and statistical issue that should be incorporated into the study protocol during the experimental set-up and analysis. This is emphasized only because, if there are issues with non-endocrine-linked endpoints that potentially confound interpretation of the endocrine-linked endpoints, then efforts should be made by the Agency through more in depth analysis of available existing data (e.g., for Chemical J) and/or protocol changes in the future to clarify the data in support of the WoE approach.

Charge 2.4. Chemical A illustrates a situation where a molecular event has been initiated along a pathway via binding to the androgen receptor and by altered steroidogenesis, with corroborative evidence from the Hershberger assay. However, at a higher level of biological organization, an anti-androgenic response is not expressed within the context of the mammalian intact hypothalamic-pituitary-gonadal axis (based on the Tier 1 mammalian assays and the mammalian in vivo OSRI). In the absence of overt toxicity, there were some possible endocrine-related responses in the FSTRA, but they lacked diagnostic specificity (e.g., reduced GSI and altered histology). The Agency concluded that although there is evidence of an endocrine interaction (i.e., the androgen signaling pathway) at lower levels of biological organization, clear endocrine-driven responses are not expressed at higher levels of biological organization in organisms with an intact HPG-axis, presumably due to compensatory processes.

Please comment on how the Agency has integrated different sources of data along a biological continuum to characterize endocrine interactions of Chemical A and the transparency and reasonableness of the decision logic.

Panel Response:

Chemical A is an OP insecticide with cholinesterase inhibition as the primary MoA. It is soluble in water, somewhat volatile, and does not appear to accumulate because it is degraded by both biotic and abiotic mechanisms. Based on the WoE for Tier 1 and OSRI data, EPA

concluded that Chemical A does not demonstrate potential to interact with the estrogen pathway or the thyroid pathway, but that Chemical A has the potential to interact with the androgen pathway.

Review of the Tier 1 estrogen pathway data indicated that the ER binding and aromatase assays were equivocal and the ERTA and steroidogenesis assays were negative. The *in vivo* tests included the uterotrophic and female pubertal assays, which were negative for estrogen-related responses. In contrast, there were some responses in the FSTRA in the absence of overt toxicity, but they lacked diagnostic specificity (e.g., effects on male gonadal tissue or GSI). Results in the FSTRA were complicated by evidence of overt toxicity, particularly at the high concentration. In the FSTRA, there were effects of decreased GSI in males at the high and moderate doses; altered testis histology (spermatogenesis) in males at the high, moderate and low doses; and follicular development in females only at the high dose, and fertility and fecundity were reduced at the high dose. The tubercle score was altered in males only at the high dose and VTG was decreased in females at the high and low dose. Additional discussion of overt toxicity in the FSTRA was discussed previously in the Panel's response to Charge 2.2.

For the androgen pathway, there was evidence of AR binding and altered steroidogenesis, resulting in reductions in testosterone. In the Hershberger assay there were lowered accessory sex organ weights in the anti-androgenic arm of this assay. None of the Hershberger assay results were affected by overt toxicity as was the case in the FSTRA. OSRI data also indicated developmental delays in rats, as there was a slight delay in preputial separation. Based on these results the EPA concluded that a MIE (i.e., AR binding and altered steroidogenesis) could be conserved across different taxa. The Tier 1 FSTRA also indicated that an anti-androgenic effect may be expressed, but at concentrations that are overtly toxic and affected by cholinergic effects. EPA further concluded that fish may be more responsive than the rat to these MIEs/AOPs (particularly altered steroidogenesis). While it is clear that fish were more sensitive to the AChE effects of Chemical A, which may affect some aspects of the pathways, it is unclear why they would be more sensitive to endocrine substances. Question 2.4 is premised on assumptions about identifiable AOPs or MOAs for the test compounds, and also on the inference that "a molecular event has been initiated via binding to the androgen receptor and by altered steroidogenesis". It is important to examine these presuppositions in order to fully comment on the specifics of the question.

In order for a MIE, such as receptor binding, to be relevant as an endocrine disruptor that drives a particular AOP, certain quantitative relationships need to be established. In the case of Chemical A, its apparent affinity is approximately 7 orders of magnitude less than methyltrienolone (R1881), which has potency on rat AR that is similar to dihydrotestosterone (DHT). The WoE information provided does not explicitly address the likely (or known) exposure levels for Chemical A. Working backward, and assuming a bioavailable testosterone concentration of about 100 ng/dl (3 nM), Chemical A exposure will need to be on the order of tens of molar (10^{-2}M) to exert biological effects via this MIE. In the context of a WoE discussion, characterization of a compound as a "binder" does not provide sufficient information to determine with any precision how much weight should be placed on this particular evidence. Since the primary pesticidal MoA for Chemical A is cholinesterase inhibition, and the compound does not appear to accumulate, one supposes that it would manifest toxicity via cholinesterase

inhibition at concentrations much below those implied for androgen pathway activity by the extremely weak apparent AR affinity. With sufficient information this could be calculated; the Panel recommended this approach.

The quantitative consideration for the AR MoA for Chemical A has important implications because it may, therefore, be possible to interpret the lack of quantitatively relevant binding as being consistent with, rather than contradictory to, the lack of evidence for any androgen-related responses in the mammalian intact pubertal models.

The evidence for activity of Chemical A in the steroidogenesis *in vitro* studies in H295R (adrenal) cells is presented in White Paper (section 6.C.v., Table 7 and Figure 10, and is summarized in Table 18) as a "decrease of testosterone" (i.e., inhibition of androgen biosynthesis). Examining the data in as much detail as is possible, the evidence for a meaningful inhibitory effect on androgen biosynthesis would have to be considered fairly weak and tentative. There is no dose-response relationship (T is flat at all doses up to 10 µM, then drops to 84% of control). In the task guidelines for the EDSP, any value that is statistically significant relative to the control is considered a positive result, hence, the need to interpret the current Chemical A results as inhibition of androgen biosynthesis. In contrast, the OECD guidelines require that two adjacent concentrations be significantly different for the test to be considered a positive result. This OECD criterion guards, to a greater degree, against spurious effects.

If one were to apply more stringent interpretations for the AR binding and steroidogenesis tests, one could surmise that these molecular and cellular level findings are fully consistent with results from the uterotrophic assay, the intact mammalian reproduction models (pubertal female and male rats), and with the FSTRA (discounting data with overt toxicity). In addition, the results are consistent with the lack of androgen agonist effects in the Hershberger assay. Therefore, no "compensatory processes" need to be invoked. The non-conforming results according to this view are the clear findings in the androgen antagonist arm of the Hershberger assays.

The dose-related anti-androgen results across the full spectrum of male secondary reproductive organs appear to be a very robust finding for Chemical A. These results are neither complementary nor redundant with other assays. The AR binding affinities are much too low to account for any direct anti-androgen receptor-binding MoA, and the steroidogenesis assay results are irrelevant because testosterone propionate (TP) is added exogenously. Consequently, the apparent antagonistic androgen effect has to remain an "unexplained" effect, which does not benefit from any reliable complementary or redundant supportive evidence in any of the other assays. Whether this result, which appears to be robust, is sufficient or not for recommending Tier 2 testing remains an unanswered question.

The case raises, in a very clear instance, the problem that the battery of assays may produce clear results that are largely unsupported by other assay battery findings. In such cases, what would be the logic for making further recommendations? Since Tier 1 is a "screening" exercise, an argument can be made that any single reliable positive finding, which the Hershberger antagonist findings appear to be, would be sufficient to move Chemical A to some next level of testing. However, an equally powerful case could be made that a single test, such

as one arm of the Hershberger, is insufficient evidence without having either a reliable mechanism of action to explain it, or other complementary evidence in different *in vivo* assays. How these choices will be made remains unclear.

For the T pathway, results in the female pubertal assay indicated slight T4 reduction, which was considered equivocal, along with no complementary changes in TSH, thyroid weight or thyroid histopathology. In the AMA there was thyroid gland hypertrophy, but only at the highest concentration. In other mammalian studies there were no changes in thyroid weights, pituitary weights or thyroid histopathology in the sub-chronic and chronic toxicity studies. Lack of complementarity and redundancy within and among (frogs and rats) bioassays, and no contradicting evidence in the OSRI indicates no effects on T. EPA, thus, concluded appropriately that there was insufficient evidence to demonstrate a potential to interact with the thyroid pathway.

Charge 2.5. In some chemical situations, the in vitro Tier 1 data are negative. Nonetheless, this does not necessarily detract from a conclusion of a potential endocrine interaction in vivo either because a different molecular initiating event (MIE) may be occurring than what the in vitro assay evaluates or because an activated metabolite may be responsible for the in vivo effects. Chemicals N and S provide an illustration of this situation in that the MIE is uncertain due to the negative Tier 1 in vitro assays. But, there were Tier 1 in vivo responses that are consistent with potential interactions with the androgen or estrogen signaling pathways.

For Chemical N, anti-androgen related responses were observed in the male pubertal assay that were complementary within the assay (i.e., decreased in testosterone levels that progressed to effects at the organ (tissue weight decreases in androgen sensitive tissues) and organism level (delay in PPS). In the FSTRA, more limited responses were observed in the absence of overt toxicity, i.e., a decrease in female VTG that did not manifest into higher level effects. In this case, there is in vivo evidence of an endocrine interaction but compared to other case studies (e.g., as Chemical J), the complementarity and redundancy in responses are not as robust.

In the case of Chemical S for the A pathway, in the Hershberger there was a decrease in androgen-sensitive tissue weights. In the case of the male pubertal assay, there were complementary responses in that a cellular response (i.e., decreases in testosterone levels) progressed to effects at the organ (tissue weight decreases in androgen sensitive tissues) and organism level (delay in PPS). In the FSTRA, there were altered male gonadal weights and reduced tubercles. Although these effects in the fish lack specificity, they are supported by the mammalian responses. Tier 1 in vivo responses are not observed at the lower concentrations in organisms with an intact HPG-axis, presumably due to compensatory processes.

Please comment on the how the Agency has integrated different sources of data along a biological continuum to characterize this endocrine interaction and the transparency and reasonableness of the conclusion drawn.

Panel Response:

The Panel agreed that negative Tier 1 *in vitro* data do not detract from or limit the conclusion of a potential endocrine interaction based on positive data observed in the *in vivo* assays only. As indicated by the Agency, it is entirely possible that an unanticipated molecular initiating event or metabolism of the chemical to an active intermediate may account for interaction with the endocrine system.

The case study presented in the White Paper for Chemical N indicates all *in vitro* tests including the ER binding assay, the ERα transcriptional activation assay, the AR binding assay, the steroidogenesis assay, and the aromatase assay were negative. Similarly, Chemical S was negative for the ER binding assay, the ERα transcriptional activation assay, the AR binding assay, and the steroidogenesis assay. The outcome for Chemical S in the aromatase assay was equivocal. Chemical S did inhibit aromatase activity and in a dose-related manner. However, maximum inhibition at 10^{-5} M did not reach the 50% guidance criteria to be considered positive; aromatase activity in the presence of Chemical S was reduced to approximately 68% of control. Thus, the effects of Chemical S on aromatase were considered to be equivocal.

In contrast to the *in vitro* assays for Chemicals N and S, Tier 1 *in vivo* responses that are consistent with potential interactions with the androgen or estrogen signaling pathways were reported.

For Chemical N, potential endocrine interactions were observed in the male pubertal assay and the FSTRA. In the male pubertal assay three endocrine related endpoints were affected. The age of PPS was delayed from a mean of 44.7 days in the control to 46.5 days in the animals treated with 800 mg/kg/day. No effect of treatment on the age of PPS was noted in the 50 and 200 mg/kg/day treatment groups. Reduced serum testosterone levels were also observed following treatment. Serum testosterone levels were decreased ($p<0.05$) by 58% in the highest dose group. Levels were also decreased in the 50 and 200 mg/kg/day groups, by 25% and 41% respectively; however, these decreases were not statistically significant. Finally, the weights (adjusted, absolute, and relative) of the seminal vesicles and coagulating glands were decreased by 24-26% in the 800 mg/kg/day dose group compared to controls. The effects of Chemical N were observed in the absence of clinical signs of toxicity. The conclusions made in the White Paper were that Chemical N was positive for the androgen pathway and negative for the thyroid pathway in the male pubertal assay. In the FSTRA, a single endpoint was altered; VTG was decreased by 44% in female fish exposed to the middle dose level, a nominal concentration of 8.0 mg a.i./L Chemical N. This was the only endpoint altered in the absence of toxicity. Based on the biological control of VTG production, this result could be interpreted as either a direct estrogen antagonism or a result of androgen agonism or antagonism.

For Chemical S potential endocrine interactions were observed in the Hershberger assay and the female pubertal assay. Chemical S exhibited anti-androgen activity at the lowest tested dose (10 mg/kg/day) where significant decreases ($p<0.05$ and $p<0.01$, respectively) were noted in the ventral prostate (decreased 19% compared to testosterone propionate) and glans penis (decreased 16% compared to testosterone propionate). These effects were observed in the absence of overt toxicity and thus, Chemical S was considered to exhibit anti-androgenic

activity. In the female pubertal assay, the age at vaginal opening was altered in rats treated with Chemical S. The age of vaginal opening, both unadjusted and adjusted was increased ($p<0.01$) by 2.1 to 2.9 days in the low dose, 10 mg/kg/day group. This effect was observed in the absence of toxicity and thus, it was concluded that Chemical S was positive for the estrogen pathway and negative for the thyroid pathway.

Regarding the WoE analysis for Chemicals N and S, significant additional OSRI was included for consideration that represented a diversity of organisms and assays.

For Chemical N, OSRI included an ERα yeast two-hybrid study assessing ligand receptor binding and an ERα chloramphenicol acetyltransferase transcriptional activation assay. In addition, two large *in vivo* studies were included: a 90-day sub-chronic oral toxicity study in male and female Sprague-Dawley rats where effects on body weight, testes, thyroid, prostate, ovaries, uterus and pituitary were examined and a 90-day sub-chronic oral toxicity study in beagle dogs where effects on testes, prostate, seminal vesicle, ovaries, uterus, pituitary and thyroid were examined. In each of these studies, endocrine-related effects of Chemical N were not observed.

Carcinogenicity of Chemical N assessed in male and female F344/N rats and B6C3F1 mice did indicate evidence of renal tubular cell adenomas and adenocarcinomas and carcinomas of the preputial gland in male rats and increased incidences of hepatocellular adenomas and carcinomas and mesenchymal tumors of the integumentary system (fibroma, fibro sarcoma, neuro fibrosarcoma, or sarcoma) in male mice. Although these studies indicated potential carcinogenicity of Chemical N, no evidence of endocrine interaction was found.

Developmental toxicity studies were carried out in female CD-1 mice and Fischer 344 rats where exposure to Chemical N occurred during gestation days 6-15. Developmental effects were not noted at the highest dose tested in mice (115 ppm) and a decreased fetal crown-rump distance in female fetuses was noted in rats at the highest dose (115 ppm).

As indicated in the White Paper, no ecotoxicity OSRI data were available for Chemical N. Additionally, ASTER was used in considering the WoE analysis for Chemical N.

OSRI considered in the WoE analysis of Chemical S included Part 158 mammalian and ecotoxicity studies previously submitted to the Agency. Studies included in the analysis and noted in the White Paper included a subchronic 3-month toxicity study in Carworth Farm E (CFE) male and female rats where histopathology of testis was assessed and a 2-year CFE rat toxicity study where testes, prostate, ovaries, uterus, mammary glands, pituitary and thyroid glands were examined. In these studies no effects of Chemical S on endocrine related endpoints were noted.

An 18-month carcinogenicity study in mice included examination of testes, ovaries, prostate, uterus, pituitary and thyroid glands and no effects on endocrine related endpoints were noted.

Two developmental toxicity studies were included in the analysis. First was a developmental toxicity study in Wistar rats where Chemical S was administered to pregnant females on days 6-15 of gestation. In this study, no developmental effects of Chemical S were noted.

In another study, female New Zealand white rabbits were administered Chemical S on days 6-18 of gestation. The LOAEL was established as 10 mg/kg bw/day based on increased numbers of resorptions/litter, post-implantation loss/litter, and decreased fetal body weights. However, significant toxic effects of the chemical were associated with these effects and thus, whether these outcomes were related to effects of Chemical S on the endocrine system could not be established.

In addition, a two-generation reproduction study carried out in rats was considered in the WoE analysis. In this study no effects on fertility were noted.

Chemical S was administered for 2 years duration to beagle dogs. Organ weights were obtained for testes and thyroid and no effects of treatment were found on these parameters.

Ecotoxicity OSRI was also considered in the WoE analysis. These studies included an early life stage test in rainbow trout and an acute toxicity study in rainbow trout. Results from each study indicated Chemical S exhibited no measurable effects on endocrine endpoints.

Two studies in birds were also considered in the analysis: a one-generation reproduction test with bobwhite quail and a 19-week study in mallard duck. Results from both studies indicated a reduction in the number of eggs laid per hen, and an increase in the number of hens that exhibited lesions of old egg yolk peritonitis at terminal sacrifice. These effects were seen at the highest treatment dose (500 ppm) in both studies.

Finally, the freshwater invertebrate waterflea, *Daphnia magna* was exposed in a 21-day chronic toxicity test where no effects were noted in the absence of toxicity. The NOAEL was 16 µg/L (measured) or 25 µg/L (nominal). There was a significant difference between the solvent control and the test chemical at the two highest treatment concentrations (50 and 100 µg/L, nominal) with respect to growth, percent adult survival, and total number of young. However, these NOAEL values are above the reported solubility of Chemical S, which is about 13 µg/L.

As noted in the White Paper, no additional data with frog or other amphibian species have been submitted to the Agency for Chemical S.

Based on the case studies presented in the White Paper for Chemicals N and S, the Agency has included a large body of OSRI representing a range of organisms and tests in the WoE analysis. Further, the Agency has integrated this information into the overall WoE analysis for Chemical N and S. The Panel found the information proved to be transparent and the conclusions drawn to be reasonable.

Charge 2.6. In each of the cases studies, there was a lack of anticipated complementary and redundant responses indicative of a chemical's interaction with the thyroid signaling pathway. In the rat, there were T4 changes that were either marginal or equivocal (Chemical A), or isolated organ weight changes (Chemicals J and S) or histopathological changes of the thyroid gland (Chemical J) that were not coincident with hormone changes. In the AMA, there were some isolated responses not necessarily indicative in terms of the endpoint specificity of a hypothalamic-pituitary-thyroid axis perturbation (Chemicals A and N). The Agency considered the lack of complementarity and redundancy in responses to support a conclusion of no interaction with the HPT axis, and viewed these isolated responses insufficient to support a conclusion of an interaction with the thyroid signaling pathway.

Please comment on the how the Agency has characterized this endocrine interaction at different levels of biological organization, and the transparency and reasonableness of the conclusion drawn.

Panel Response:

Based on the limited repertoire of *in vivo* HPT axis assays in the Tier 1 battery, the Agency concluded correctly that there is insufficient data to support a conclusion that any of the test compounds interact with the thyroid-signaling pathway. The general lack of effects on thyroid hormone production for any of the test compounds does not support moving any to Tier 2 for HPT axis evaluation. Because no *in vitro* assays exist in the battery to evaluate either molecular or cellular interactions, the assays in the battery mainly evaluate factors that affect plasma thyroid hormone concentration (which may reflect hormone biosynthesis/secretion/metabolism/clearance) and thyroid gland structure (histopathology, which may mirror changes in plasma thyroid hormone). The battery of assays cannot distinguish effects on biosynthesis vs. secretion vs. metabolism vs. hormone action. Failure to affect plasma TSH concentration in rats, or the rate of tadpole development supports lack of effect on thyroid hormone receptor-dependent signaling. However, the Tier 1 thyroid screening battery is incomplete, and it is unclear whether the assays will allow for assessment of all thyroid hormone receptor (TR)-dependent processes, or signaling that does not depend on the nuclear receptors.

Like the E and A pathways, the complexity of the HPT axis has grown with the realization that both canonical and non-canonical MoAs contribute to the biological outcome especially at the organism level. With respect to both classical and non-classical MoAs, the potential to interact and alter the HPT axis occurs at each segment of the feedback loop (Cheng, et al., 2012 and Lin et al., 2012).

For example, changes in TSH and rates of tadpole limb differentiation (and other processes occurring at metamorphosis) are TRβ-dependent. TRα-dependent processes may or may not be captured in the assays (the only TRα-dependent process is initial tadpole limb outgrowth – but not digit differentiation which is TRβ-dependent). Thyroid-active chemicals may differently impact these receptors. In rodents, the TRα isoform mediates hormone dependent gene expression in many tissues and selective agonists exist that distinguish between the two TRα and TRβ isoforms. The likelihood that test compounds may show similar discriminating interactions confounds analysis of classical HPT axis in whole animals.

Importantly, test compound-dependent alterations in the binding of thyroid hormone to serum proteins is a common event that will alter the HPT axis in the short term and result in a resetting of the axis. Serum binding (more that 99.9% of the thyroid hormone is bound to serum proteins) is further confounded by differences in the cohort of serum proteins that bind T4, T3 and other TH metabolites across species. For example, in rats, a number of physiological perturbations, caloric restriction, and altered liver and/or kidney function impact the relative concentrations of both T4 and T3 in blood by altering the abundance of thyroid hormone binding proteins. Rats normally lack thyroid-binding globulin (TBG), the T4 binding protein in humans, but express this potent serum binding protein under a number of physiological conditions, especially caloric restriction. This leads to a significant increase in total circulating T4 concentration with no compensatory changes in TSH or thyroid histology because the "bioavailable" free thyroid hormone levels remain unchanged. In mammals, and also humans, generic "illness" leads to the euthyroid sick syndrome characterized by normal to elevated T4, depressed T3 and no change in TSH. This widely recognized syndrome violates the canonical HPT axis relationship – high serum thyroid hormone suppresses TSH and "quiets" changes in thyroid morphology.

Thus, the current Tier 1 assays for the HPT axis cannot provide the degree of confidence for screening that is available for the E and A axis. For screening purposes, TSH and T4 measurement is used by clinicians world-wide to assess thyroid status and it remains the gold standard in vertebrate studies.

Charge 2.7. In the absence of Tier 1 data, OSRI was available that indicated effects on thyroid endpoints in the rat (Chemical X) but the results were inconsistent within and among studies and there was no OSRI presented from amphibian studies.

Please comment on the how the Agency has characterized this endocrine interaction at different levels of biological organization, and the transparency and reasonableness of the conclusion drawn.

Panel Response:

The OSRI provided for Chemical X is not sufficient to exclude the HPT axis as a potential target and further study is required. As in the case detailed in the response to Charge 2.6, the multiple sites for potential xenobiotic interactions along the HPT axis require additional studies to evaluate interactions between Chemical X and the HPT axis. Based on the limited OSRI dataset, the Panel concluded the Agency appropriately evaluated the potential for Chemical X to impact the HPT axis.

Charge 3.0. Based on all of the case study analyses, please provide overall comments on how the Agency has employed its WoE guidance and characterized the evidence and conclusions and include in your response the following points:

a. How consistent and transparent the cases studies are in terms of documentation.

Panel Response:

In general, these data are as consistent and transparent as the bioassay results allow for each of the compounds tested. No two compounds had similar results and what has been observed is a consistent, open and honest attempt by EPA to evaluate and integrate the findings consistently for each compound. For the most part, the Panel found the documentation of the case studies consistent and transparent; however, inconsistencies were noted in documenting how overt toxicity and significance of endpoints was implemented across chemicals. For example, overt toxicity at the medium concentration for Chemical A and treatment of decreasing VTG levels appear to be treated differently for Chemicals A and N.

It was not transparent why control performance was deemed acceptable in some examples when several validity and performance criteria were not met. For example, with Chemical A, fecundity was less than the 15 eggs/female/day/replicate and spawning frequency was greater than every four days for two of the four control replicates, and fertilization success was less than 95% in control. Dissolved oxygen also dropped below the required minimum percent saturation of 60%. The White Paper states the reason for including it as "Although some control performance measures were less than recommended, in general, the performance was reasonable." Public commenters pointed out that the apparent inability of Tier I performers to obtain quality control criteria could result in unreliable data. Therefore, a more cogent explanation is required to understand why the control performance is still considered valid even though several criteria were not met.

Transparency would be greatly facilitated if mode of action for related and characterized chemicals was also considered as a read across or as classes of chemicals (e.g., chemicals that impact fish but not the rat assays).

As mentioned under Charge 1, using the same format to present the data for each case was helpful.

The OSRI approach for Chemical X provided a contrast from the other Tier 1 battery assessments for the other compounds by having more independent replication testing for some test endpoints. The reporting/quantification of this independent replication for certain bioassays provided in the summary tables provided greater transparency of the data and also made sure that results between the same bioassay were consistent.

One Panel member thought that as more data for specific AOPs/MIEs are developed, consistency and transparency will continue to improve.

Public comments indicated a need for EPA to better define how results from Tier 1 will ultimately be used to determine the need for Tier 2 Testing versus requiring additional screening data under Tier 1, such as better dose spacing to better define possible endocrine disrupting chemical effects.

The implementation of WoE, as practiced by the EPA, appears to be an integrative and qualitative process based on professional judgment. As demonstrated by the case studies and the public comments, Tier I data appear to be more difficult to integrate due to the very different data types in the absence of a quantitative weighting. Therefore, a more explicit definition of implementation of assay/endpoint weighting should be acknowledged in the WoE analysis. This need is reflected in both the public and the Panel's comments where the development of a more rigorous quantitative approach is desired and a potential framework for implementation presented. The framework presented by the Endocrine Policy Forum provides an example of a clear and transferable approach for integrating data from multiple scales and multiple levels of biological organization. Such an explicit acknowledgement of evidence weighting would greatly facilitate interpretation, transparency and consistency of Tier 1 and OSRI data across chemicals. Some weighting is noted in the White Paper, but was contradicted in oral comments by the Agency.

As indicated by the public commenters, there are differences in what the EPA presented in the White Paper and the statistical analyses and conclusions drawn from the same data by the registrants. If the Agency now has reasons to change the analyses that are part of the guidelines, as part of the iterative process of developing and interpreting the Tier 1 battery, the Agency should make these changes explicit in the guidelines.

There was consensus on the Panel that the overall approach need not necessarily be highly quantitative for a screening level set of tests, but the use of a systematic and transparent decision tree would be of great use to everyone. The Panel commended the Agency on the use of the FSTRA decision tree and it appears there is consistent decision logic being applied to make judgments on the case study chemicals. Currently, these decision rules are not entirely transparent and the Panel recommended approaches for integrating data in the fashion as highlighted by public commenters. When this is not possible though, the addition of key descriptors of each of the endpoints and how it integrates in evaluating the potential for E, A, or T pathways would be helpful for both transparency and consistency in decisions made by the Agency. In particular, written guidance on a qualitative indication of how important each relative endpoint is, and how correspondence (or lack of correspondence) between various endpoints provided guidance into particular decisions would be helpful. Without the presence of the decision trees, which might be difficult to create for all the Tier 1 tests, a clear description of how and why the decisions were made to indicate the potential (or lack of potential) for interaction along these pathways should be included in this integrative fashion for the discussion of chemicals.

b. How adequately the Agency has described the extent of complementarity and redundancy of responses and has integrated and interpreted diverse lines of evidence across different biological levels of organization and taxa to reach preliminary conclusions regarding endocrine interactions.

Panel Response:

The results of the 2008 SAP presented by EPA was helpful to better understanding how the Agency rates findings across individual bioassays and across batteries of bioassays. This provided greater insight into the process for data integration within the Tier 1 assessment. More visualization of this process, such as by presenting converging AOPs, would help make this a more understandable and transparent process. The decision logic trees for the FSTRA and the AMA were helpful and provided clear logic for how data results were interpreted. Similar decision trees are needed for each of the E/E-, A/A-, Steroid Synthesis T/E; HPG and HPT assays and pathways and how the batteries would fit into the logic.

While the qualitative approach using best professional judgment is adequate for now, the process should move towards a more quantitative approach as more fundamental findings from high throughput ToxCast and Tier I screening program are gleaned. The ranking approach suggested by some of the public commenters has consistency with the concepts of complementarity within an assay and redundancy among different bioassays. Quantification of an integrative process provides greater clarity than qualification of an integrative process as the WoE provides currently.

The summary slides presented by EPA, which inserted individual bioassays along an AOP for the E, A and T signaling AOP pathways, were very helpful and improved the clarity of issues of redundancy and complementarity. Adding additional summary tables to these figures, which outline at a high level the decision-making process across each E, A, T pathways, was very helpful and greatly enhanced the ability to interpret integration within and among signaling pathways.

Overall, the Agency has adequately described the extent of complementarity and redundancy of responses and the Panel provided comments on complementarity and redundancy for Chemical J in its response to Charge 2.3. Also in the Panel's response to Charge 2.3 they also pointed out where there are reasonable scientific reasons to accept lack of complementarity and redundancy (e.g., Chemical J, effects on steroidogenesis).

It was unclear how different results from different taxa were treated. Species-specific responses do not appear to be given a significant weight in analysis. Since taxa might respond differently, there might not necessarily be redundancy or complementarity of some effects. How this is treated or considered in the overall evaluation of potential for endocrine interactions is not adequately described. Chemical N provided examples where a strong endpoint near the level of overt toxicity existed in the fish assay. If overt toxicity exists in fish at a high concentration and only a few indicators of endocrine interactions occur at the lower concentration with no corresponding endpoints in mammalian tests or *in vitro* tests, would this be adequately considered as having potential? The example with Chemical N appears somewhat similar to

effects seen in fenarimol where an effect is seen in the FSTRA, but not in the rat pubertal assay (Ankley 2005).

c. How the Agency has used OSRI data to further characterize the observations from EDSP Tier 1 assays in determining potential chemical interactions with the E, A, and T signaling pathways.

Panel Response:

OSRI was helpful in discerning overt toxicity for Chemical A. ASTER and related databases could be of great assistance as a pre-screen for Tier 1 Testing. The OSRI-based example for Chemical X shows the power of this approach and, where there are independent repeated bioassays, the ability to include an assessment consisting of the response in WoE summary findings, which is helpful.

As noted previously in the Panel's response to Charge 1.1 (e.g., developmental effects in a mysid test for Chemical J), the current framework does not address how it will use OSRI data that falls outside of results for the E, A and T assessments made in the current Tier 1 screen. A more defined process for adding or removing data for inclusion/exclusion with OSRI and the use of the current Data Evaluation Record process used by EPA, which may include other data on non-endocrine effects as well as endocrine effects as suggested in public comments, would be very advantageous.

In the examples given, the Agency has adequately used OSRI data to further characterize the observations from the EDSP Tier 1 assays. Inclusion of Chemical X, a weight-of-evidence approach using OSRI only, adds to the argument that OSRI can provide significant value to the Tier 1 battery approach. It is helpful that the EPA provided the Panel with extra information during the SAP process on how OSRI are assessed to contain "good data". It will always be challenging for the EPA to ensure they have captured all the relevant external data; the EPA will need to remain diligent in its approach to being aware of and undertaking assessment of all available relevant data. The EPA should consider further discussion around the issue of what OSRI is actually "relevant". For example, when are invertebrate data relevant? When are they not? How close to the Tier 1 battery do studies on vertebrates need to be to be considered "relevant"?

Toxicological data from related chemicals could provide more support for E, A and T pathway impacts, especially where effects in one species are not supported by corresponding effects in another. The Agency's WoE approach would also greatly benefit from considering data and models from high throughput screening efforts such as ToxCast in evaluating the OSRI.

d. How the Agency has considered the understanding of a chemical's mode of action and how that informs the weight that is placed on Tier 1 responses in the presence of uncertainties introduced by dose setting, overt toxicity, and portal of entry issues.

Panel Response:

The Panel concluded that overt toxicity is a significant issue, as demonstrated in the FSTRA for Chemical A, and they recommended that the Agency design a more precise or prescriptive approach for consistently determining effects of overt toxicity, particularly in determining the lower limits.

ASTER and other toxicity databases should be fully utilized as a Tier 1 prescreen for as many chemicals as possible. The case study for Chemical X based on OSRI only was a very good indicator of the utility of this approach.

The log-based dose design of some Tier 1 assays limits the ability of those assays to produce consistency in terms of dose-response relationships. Review of the shape of the dose-response curves for existing acute and chronic toxicity data would at least provide better guidance on dose selection for those compounds with steep slopes in the dose-response curve. Another aspect of dose not discussed much was the potential difference in internal dosing resulting from oral dosing in mammalian studies and the aquatic exposure routes in the FSTRA and AMA. In mammals, exposure is via oral route (digestive tract) under acidic conditions where as in the FSTRA, it is via the gill (respiratory organ) at generally a neutral pH. In the AMA, frogs are exposed aquatically under generally neutral pH conditions. The variations in internal doses resulting from these exposure routes may be profoundly different, and may affect results more so than log spaced results. Therefore, exposure route (portal of entry) may be a critical consideration (e.g., oral dosing compared to waterborne or injection exposure includes first-pass metabolism to the liver) that may affect toxicity.

Dimensionality and magnitude of effects help determine the certainty and consistency of results leading to robustness, which EPA demonstrated with several examples in the case studies.

Issues raised during previous Panel comments included a need for the Agency to consider dose setting as it pertains to overt toxicity (i.e., chemicals need to be tested at a level which does not elicit overt toxicity responses). For example, in the FSTRA, because of the log spacing of test concentrations, there is ample room between the highest and medium concentrations to allow testing of a relevant concentration between them if the highest concentration causes overt toxicity.

In cases where overt toxicity is observed, the Agency should refrain from making conclusions on endocrine potential until the overt toxicity issue is resolved. This will protect from concluding that there is the potential for endocrine effects when there is none or concluding that there is no potential for endocrine effects when potential exists.

In general, for the case studies provided, the Agency has considered these key issues to some extent and should continue, in an iterative fashion, to develop ongoing guidance to deal with these factors as more data are analyzed. When overt toxicity levels are not considered or when entire assays are not included due to solubility issues, as with Chemical S, it is not clear how or if the WoE approach is adjusted in any way to deal with the reduction in size of the overall battery.

REFERENCES

Ankley G.T., Jensen, K.M., Durhan, E.J., Makynen, E.A., Butterworth, B.C., Kahl, M.D., Villeneuve, D.L., Linnum, A., Gray, L.E., Cardon, M., and Wilson, V.S. 2005. Effects of two fungicides with multiple modes of action on reproductive endocrine function in the fathead minnow (*Pimephales promelas*). Toxicol Sci. Aug;86(2):300-8. Epub 2005 May 18.

Ankley, G.T. and Gray, L.E. 2013. Cross-species conservation of endocrine pathways: a critical analysis of Tier 1 fish and rat screening assays with 12 model chemicals. Environ. Toxicol. Chem. 32:1084-87.

Cheng, S.Y., Leonard, J.L., and Davis, P.J. 2010. Molecular aspects of thyroid hormone therapy. Endocrine Reviews. 31:139-170.

Fulton, M. H. 1989. An assessment of the sublethal effects of selected organophosphate insecticides in laboratory and field exposures with the mummichog, *Fundulus heteroclitus*. Univ. Of South Carolina, Arnold School of Public Health, Ph.D. Dissertation: 110 pp.

Fulton, M.H. and Scott, G.I. 1991. The Effects of certain intrinsic and extrinsic variables on the acute toxicity of selected organophosphorus insecticides to the mummichog, *Fundulus heteroclitus*. Journal of Env. Sci and Health B26:459-478.

Kavlock, R.J., Daston, G. P., DeRosa, C., Fenner-Crisp, P., Gray, L. E., Kaattari, S. Lucier, G., Luster, M., Mac, M. J., Maczka, C., Miller, R., Moore, J., Rolland, R., Scott, G., Sheehan, D. M., Sinks, T., and Tilson, H. A. 1996. Research Needs for the Risk Assessment of Health and Environmental Effects of Endocrine Disruptors: A Report of the U.S. EPA-sponsored Workshop, Environmental Health Perspectives 104 (4): 1-26.

Lin, H.Y., Tang, H.Y., Davis, F.B., Mousa, S.A., Incerpi, S., Luidens, M.K., Meng, R., and Davis, P.J. 2012. Nongenomic regulation by thyroid hormone of plasma membrane ion and small molecule pumps. Discov. Med. 14(76):199-206.

Van Dolah, R.F., Maier, P.P., Fulton, M. H. and Scott, G.I. 1997. Comparison of azinphosmethyl toxicity on juvenile red drum, *Sciaenops ocellatus*, and the mummichog, *Fundulus heteroclitus*. Env. Toxicol. and Chem. 16:1488-1493.