

Original Article

Relevance Weighting of Tier 1 Endocrine Screening Endpoints by Rank Order

Christopher J. Borgert,^{1*} Leah D. Stuchal,² Ellen M. Mihaich,³ Richard A. Becker,⁴ Karin S. Bentley,⁵ John M. Brausch,⁶ Katie Coady,⁷ David R. Geter,⁸ Elliot Gordon,⁹ Patrick D. Guiney,¹⁰ Frederick Hess,⁶ Catherine M. Holmes,⁶ Matthew J. LeBaron,⁷ Steve Levine,¹¹ Sue Marty,⁷ Sandeep Mukhi,¹² Barbara H. Neal,¹³ Lisa S. Ortego,⁸ David A. Saltmiras,¹¹ Suzanne Snajdr,⁵ Jane Staveley,¹⁴ and Abraham Tobia^{15†}

¹Applied Pharmacology and Toxicology, Inc., Gainesville, Florida

²Department of Physiological Sciences, C.E.H.T., University of Florida, Gainesville, Florida

³Environmental and Regulatory Resources (ER²), Durham, North Carolina

⁴American Chemistry Council, Washington, DC

⁵E.I. du Pont de Nemours, Wilmington, Delaware

⁶BASF Corporation, Research Triangle Park, North Carolina

⁷The Dow Chemical Company, Midland, Michigan

⁸Bayer CropScience LP, Research Triangle Park, North Carolina

⁹E.Gordon Consulting LLC, Princeton Junction, New Jersey

¹⁰S.C. Johnson & Son, Inc., Racine, Wisconsin

¹¹Monsanto Company, St. Louis, Missouri

¹²FMC Agricultural Products, Ewing, New Jersey

¹³Exponent, Alexandria, Virginia

¹⁴Exponent, Cary, North Carolina

¹⁵U.S Food and Drug Administration, Silver Springs, Maryland

Weight of evidence (WoE) approaches are recommended for interpreting various toxicological data, but few systematic and transparent procedures exist. A hypothesis-based WoE framework was recently published focusing on the U.S. EPA's Tier 1 Endocrine Screening Battery (ESB) as an example. The framework recommends weighting each experimental endpoint according to its relevance for deciding eight hypotheses addressed by the ESB. Here we present detailed rationale for weighting the ESB endpoints according to three rank ordered categories and an interpretive process for using the rankings to reach WoE determinations. Rank 1 was assigned to in vivo endpoints that characterize the fundamental physiological actions for androgen, estrogen, and thyroid activities. Rank 1 endpoints are specific and sensitive for the hypothesis, interpretable without ancillary data, and rarely confounded by artifacts or nonspecific activity. Rank 2 endpoints are specific and interpretable for the hypothesis but less informative than Rank 1, often due to oversensitivity, inclusion of narrowly context-dependent components of the hormonal system (e.g., in vitro endpoints), or confounding by nonspecific activity. Rank 3 endpoints are relevant for the hypothesis but only corroborative of Ranks 1 and 2 endpoints. Rank 3 includes many apical in vivo endpoints that can be affected by systemic toxicity and nonhormonal activity. Although these relevance weight rankings (W_{REL}) necessarily involve professional judgment, their a priori derivation enhances transparency and renders WoE determinations amenable to methodological scrutiny according to basic scientific premises, characteristics that cannot be assured by processes in which the rationale for decisions is provided post hoc. *Birth Defects Res (Part B)* 101:90–113, 2014. © 2014 Wiley Periodicals, Inc.

Key words: *endocrine disruptors; weight of evidence; endocrine screening; regulatory toxicology; Endocrine Disruptor Screening Program*

Abbreviations: ADME, absorption, distribution, metabolism, and excretion; AMA, amphibian metamorphosis assay; AR, androgen receptor; ARBA, androgen receptor binding assay; CYP, cytochrome P450 enzyme; DHT, dihydrotestosterone; EDSP, Endocrine Disruptor Screening Program; EPF, Endocrine Policy Forum; ER, estrogen receptor; ERBA, estrogen receptor binding assay; ERTA, estrogen receptor transcriptional activation assay; ESB, Endocrine Screening Battery; FSTRA, fish short-term reproduction assay; GSI, gonad-somatic index; HPG, hypothalamic-pituitary-gonad; HPT, hypothalamic-pituitary-thyroid; Kd, dissociation constant; LABC, levator ani-bulbocavernosus muscle complex; LH, luteinizing hormone; MoA, mode of action; MTD, maximum tolerated dose; PPS, preputial separation; T₃, 3,5,3'-triiodothyronine, the active thyroid hormone; T₄, thyroid prohormone; the Panel, expert panel of scientists from the Endocrine Policy Forum; TR, thyroid receptor; TSH, thyroid-stimulating hormone; VTG, vitellogenin; WoE, weight of evidence; W_{REL}, relevance weight; W_{RES}, response weight

[†]The opinions and information in this article are those of the author's past work and do not necessarily reflect the views and policies of the U.S. Food and Drug Administration

*Correspondence to: Dr. Christopher Borgert, 2250 NW 24th Ave, Gainesville, FL 32605. E-mail: cjborgert@apt-pharmatox.com

Received 4 November 2013; Accepted 30 December 2013

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/dbdr) DOI: 10.1002/dbdr.21096

INTRODUCTION

Regulatory agencies have consistently agreed that endocrine disruption must be evaluated by weight of evidence (WoE) procedures (U.S. EPA, 2011a; Organization for Economic Cooperation and Development [OECD], 2012). An objective, transparent framework for hypothesis-based WoE evaluations of endocrine screening and testing data has been published (Borgert et al., 2011a), referred to hereafter as “the Framework.” Although the focus of that publication was evaluation of data from Tier 1 of the U.S. EPA’s Endocrine Disruptor Screening Program (EDSP), the proposed approach and principles apply to all toxicological and pharmacological data irrespective of the purpose or program for which data were generated.

Within the context of the U.S. EPA’s Tier 1 EDSP, a broad array of data are to be considered and several separate WoE evaluations are implied (Borgert et al., 2011a). For each WoE evaluation, the Framework recommends that “weight” and “evidence” be clearly defined to enhance transparency, consistency, and credibility. Scientific “evidence” has been defined (Gori, 2010) and is evaluated by the Framework according to primary, secondary, and tertiary validity of the data. Primary validity pertains to the soundness and quality of the measurements made in a scientific investigation. Secondary validity pertains to the thoroughness and transparency of reporting. Tertiary validity addresses the probative power of the study design and its relevance to the questions posed. These concepts are well established and when combined with recommendations for transparent reporting of literature search and selection procedures similar to those used for systematic clinical reviews, can be used to evaluate all toxicological studies (human epidemiology results, laboratory animal investigations, and mechanistic studies) used in regulatory decision making (McCarty et al., 2012). It is important to keep in mind that in the context of the Tier 1 Endocrine Screening Battery (ESB), “evidence” pertains to the potential of a substance to operate via specific modes of action (MoAs); it does not pertain to or predict an ability to cause any particular type of effect, including adverse effects.

“Weight,” on the other hand, implies that all data do not contribute equally to answering the question posed. Thus, “weighting” involves a careful consideration of the specific hypothesis to be evaluated and how each particular measurement (data) informs that hypothesis. Ideally, weight would be assigned quantitatively (Borgert et al., 2011a) based on objective measurements of predictive power, false-positive and false-negative detection rates, and potency or strength of the response. This would avoid the biases inherent in professional judgments. In practice, however, quantitative rankings are not possible because the predictive capacity of most toxicological assays for potential adverse human health effects is unknown, and this is particularly the case for endocrine-mediated toxicity. Therefore, qualitative rankings are necessary and appropriate, acknowledging that some reliance on professional judgment is unavoidable. Nonetheless, objectivity and transparency are the overriding goals.

To be of practical utility and fully transparent, two questions need to be satisfied by a WoE methodology: (1) Is sufficient methodological detail provided so that inde-

pendent analysts can apply the WoE evaluation method reliably and consistently? and (2) if independent analysts were to apply the WoE method and reach different conclusions, is the methodological detail sufficiently clear to allow a determination of why different conclusions were reached, for example, did different conclusions result from different applications of the WoE method or from different interpretations of the available data based on the same application of the method? According to McCarty et al. (2012), two factors will enable a WoE methodology to satisfy those questions: (1) the process used to weight various types of data, including its literature basis, must be clearly articulated; and (2) the weightings themselves must be derived a priori and applied consistently. The Framework employs two measures of “weight.” The relevance of each endpoint is assigned a weight, W_{REL} , according to its importance for evaluating a specific hypothesis. The strength of response produced by the test chemical in a particular assay or endpoint is also given weight, a value deemed the response weight, “ W_{RES} ” (Borgert et al., 2011a).

This article addresses the assignment of relevance weighting values (W_{RELS}) and reports the recommendations of an expert panel of scientists from the Endocrine Policy Forum (EPF) (the Panel) convened to derive a consensus rank ordering of endpoints for evaluating each of the eight hypotheses (Borgert et al., 2011a) addressed by the U.S. EPA’s Tier 1 ESB. The Panel’s recommendations are their collective scientific judgment as informed by their experience with the assays and the relevant scientific literature cited herein. These recommendations represent a transparent set of assumptions that can be used to conduct a WoE evaluation for any particular chemical, and which will produce clear and consistent WoE evaluations for substances subjected to the Tier 1 EDSP ESB.

These proposed rankings are not envisioned as the final word on weighting of the ESB endpoints for endocrine screening or as a set of definitive rankings for endocrine endpoints. Rather, it is hoped that these recommendations will encourage broader and deeper scientific discussion about the relative strengths of the specific endpoints used for evaluating and interpreting potential endocrine activity. These proposed rankings may serve as a useful starting point for a full evaluation of Tier 1 battery performance based on data from the initial round of screening under Tier 1 Test Orders (for a review, see Borgert et al., 2011b).

DEVELOPMENT OF RELEVANCE WEIGHT (W_{REL}) RANKS

The consensus “Relevance Weights” (W_{REL}) for endpoints evaluated in the 11 Tier 1 endocrine screening assays were assigned by rank ordering for each of the eight hypotheses, as summarized above. These are listed in Tables 1 through 8 as rankings 1, 2, or 3 by hypothesis, with Rank 1 being the most relevant and Rank 3 the least relevant, to the specific hypothesis. Endpoints not listed for a particular hypothesis were considered not relevant to its evaluation. In determining the W_{REL} rankings, the Panel developed several principles and the following definitions:

Rank 1: The endpoints are specific and sensitive for the hypothesis, are interpretable without knowing the response of other endpoints, and are in vivo measurements rarely confounded by artifacts or nonspecific activity.

Rank 2: The endpoints are specific and sensitive for the hypothesis, are interpretable without knowing the response of other endpoints, but are less informative than Rank 1, often due to potential confounding influences; includes in vitro and in vivo endpoints.

Rank 3: The endpoints are relevant for the hypothesis, but only when corroborative of Ranks 1 and 2 endpoints; includes some in vitro and many apical in vivo endpoints.

In the definitions above, “interpretable” means that the results for an endpoint provide information relevant to the hypothesis, without clarification from other endpoints. However, consistent with EPA policy (U.S. EPA, 2011a), a fundamental principle and requirement of the Framework is that no hypothesis can be decided on the results of a single assay (Borgert et al., 2011a). Whether a hypothesis is supported requires consideration of results from all relevant assays and endpoints (Ranks 1, 2, and 3).

Strength of Response

If an assay or endpoint is specific for a particular MoA (stated as hypotheses in the Framework), then strength of response is a reflection of the potential potency of the test substance for that MoA. In fact, whether a substance is capable of interacting with the endocrine system via any particular hormonal MoA is dependent upon its ability to exhibit sufficient potency in vivo (Borgert et al., 2013). Thus, the relevance of an endpoint for deciding any particular hypothesis, that is, its W_{REL} , depends to some extent on its ability to provide information on strength of response. The framework assigns the strength of response a specific weighting (W_{RES}), but regardless of the means by which it is considered, the strength of the response should be taken into account wherever possible. Some assays in Tier 1, such as the estrogen receptor transcriptional activation assay (ERTA), provide a means of assessing relative potencies by comparison to a positive control, and all assays allow a comparison of the dose required to elicit a response with known circulating levels of endogenous ligands or endocrine-active pharmaceuticals (Borgert et al., 2012).

In Vivo versus In Vitro Endpoints

In vitro bioassays typically detect upstream events within hormonal pathways and exhibit a high degree of specificity and sensitivity for detecting interactions with specific components of the pathway, for example, interactions of the test article with a receptor or enzyme. Such in vitro assays can provide “potency” data, based on binding affinity, enzyme inhibition kinetics, or similar measures. In vitro assays cannot, however, reliably predict the overall biological effects (i.e., downstream events) induced by a compound in vivo (Bovee and Pikkemaat, 2009). In vitro assays are also, in most cases, deliberately over-responsive (compared with many in vivo systems) toward chemicals that bind to a particular receptor (OECD, 2012).

Whereas in vivo assays allow for the evaluation of the total potential response resulting from direct and indirect mechanisms of hormonal action, in vitro assays allow for the evaluations of only those responses occurring through an individual type of system under study and do not account for in vivo processes such as absorption, distribution, binding to serum proteins, metabolism, elimination, and other pharmacokinetic processes (Gray et al., 2004; Rozman et al., 2006; Marty et al., 2011; Vitale et al. 2012). Although in vitro assays are aimed at mechanistic specificity, they can nonetheless be confounded by cytotoxicity, solubility limits, oxidative stress that compromises cell membrane integrity or transport, etc. These limitations of in vitro assays can produce false-negative results for substances that require metabolic activation, and false-positive results for both low potency substances tested at high concentrations and for compounds that undergo rapid metabolism to inactive moieties (Gray et al., 2004). Of note, high concentrations in in vitro assays are typically required in EDSP Tier 1 screening protocols, consistent with the intent to minimize false negatives at the expense of false positives within the ESB. For these reasons, Rank 1 includes only in vivo endpoints.

Ambiguity and Confounding Influences in Specific Assays that May Affect Several Hypotheses

Ambiguities in assay interpretation that affect only specific hypotheses are mentioned under the appropriate headings, but some general ambiguities and potential confounding influences are broadly applicable. These are discussed here for efficiency.

The male (EPA 890.1500) and female (EPA 890.1450) pubertal assays can detect a variety of hormonal mechanisms and may be particularly useful in cases where the hypothalamic–pituitary–gonadal (HPG) or hypothalamic–pituitary–thyroidal (HPT) axes are affected (Ankley and Gray, 2013). However, the endpoints measured in these assays are not mechanistically specific and so there is ambiguity in their interpretation. This ambiguity is exacerbated by the potential for confounding by systemic toxicity. For example, many of the endpoints can be influenced by reduced body weight gain, which is likely to occur because the EPA test guidelines specify that the highest dose level should be at or just below the maximum tolerated dose (MTD), up to a limit of 1000 mg/kg/day. These factors may explain why EPA validation studies for the pubertal assays were unable to demonstrate a negative response using a true negative control agent. Thus, it is unclear how often these assays will identify non-endocrine-active chemicals as positive based on systemic toxicity (Borgert et al., 2011b; Marty et al., 2011). Moreover, some of the endpoints measured in the male pubertal assay may not respond consistently for the same test chemical (Marty et al., 1999, 2011; Stoker and Zorrilla, 2010). The biological significance of small changes can be difficult to interpret for several endpoints evaluated in the pubertal assays, and it is unclear to what extent some changes depend on body weights (Ashby and Lefevre, 2000; Marty et al., 2003; Stoker and Zorrilla, 2010), making selection of a maximal tolerated change difficult to establish. Furthermore, several of the endpoints evaluated

in the pubertal assays are observational (e.g., histopathological evaluations, vaginal opening (VO), preputial separation (PPS)) and therefore, not as amenable to quantitative assessment as are organ weights, receptor binding and activation, hormone levels, etc. However, some organ weights, such as uterus weights in the female pubertal assay, are confounded by the onset of estrous cyclicity and are too variable to be interpreted reliably. This limits the ability of the assay to provide information on the strength of response.

The androgen receptor binding assay (ARBA; EPA 890.1150) and estrogen receptor binding assay (ERBA; EPA 890.1250) provide specific, quantitative information regarding the affinity of a test substance for a specific receptor. Although this information is quite useful for evaluating affinity, it provides no information about whether a substance might act as an agonist, partial agonist, or antagonist because it provides no information about efficacy (Borgert et al., 2013). This limits the ability of these assays to inform specific hypotheses or to provide information about the strength of the response in a particular direction. Nonetheless, the strength of the molecular interaction between test substance and the specific receptor—that is, its affinity—can assist the interpretation of results from other assays. False-positive responses in these assays can be observed for test articles that have physical-chemical properties that denature the binding site of the receptors.

Systemic toxicity can confound endpoints in several Tier 1 screening assays and needs to be differentiated in the overall interpretation of a test chemical's potential for specific endocrine activity (Marty et al., 2011). Signs of decreased body weight, decreased survival, clinical observations, and abnormal behavior indicate overt toxicity. Altered histopathology in liver, kidney, or other tissues, and alterations in certain biochemical biomarkers are also informative regarding systemic toxicity of the test material.

As discussed above, the EPA test guidelines for the pubertal assays specify a 10% body weight decrement as indicating that an MTD has been reached. However, decreased body weight in a 6 to 10% range may also affect endpoints typically considered to be potential signs of endocrine activity. Whereas an MTD based on 10% body weight reduction was recommended for the female pubertal assay, an MTD based on 6% body weight reduction may be necessary in the male pubertal assay to preclude potential effects on the levels of active thyroid hormone 3,5,3'-triiodothyronine (T_3) and prohormone thyroxine (T_4) (Laws et al., 2007). Furthermore, Marty et al. (2003) showed a 1.8 day delay in PPS from feeding restriction that produced only 10% body weight decrement, and Chapin et al. (1993) reported decreased accessory sex organ weights resulting from a 10% difference in body weights between feed restricted and control animals. Altered nutrition may also affect reproductive maturation, reproductive function, and lactation.

EPA indicates that the MTD criteria in the pubertal assays may also be reached when hepatic or renal toxicity occurs, with or without altered clinical chemistry parameters. Hepatic toxicity and associated liver enzyme induction is an important confounder that can lead to alteration in clearance of thyroid hormones and testosterone, resulting in compensatory homeostatic changes.

In the male rat pubertal assay, for example, adaptive responses can be mistaken as antiandrogenic effects. The large bolus doses administered for 30 days from postnatal day (PND) 23 to 53 can induce hepatic enzymes and produce hepatocellular hypertrophy and increased liver weights, which in turn increases testosterone metabolism (both hydration and conjugation). Except for conversion to dihydrotestosterone (DHT), testosterone metabolism markedly decreases or inactivates its androgenic activity, thereby producing an apparent antiandrogenic response that has no direct endocrine mechanism. Similar to the well-characterized MoA for enhanced metabolism and clearance of T_4 due to altered hepatic metabolism, which perturbs the HPT axis in rats, humans are expected to be markedly less sensitive to enhanced metabolism of testosterone than rats. Whereas chemical exposures can induce hepatic enzymes in both humans and rodents, rodents rely heavily on albumin to transport steroids in blood, whereas sex hormone binding globulin (SHBG) transports steroids in human blood. The dissociation constant (K_d) for steroids from albumin is 10^{-6} – 10^{-4} M, whereas the K_d from SHBG is 10^{-10} – 10^{-8} M (Baker, 2002). The low affinity binding of albumin allows steroids to more easily dissociate and undergo metabolism. Thus, the relevance of this MoA is questionable due to substantive species differences in sensitivity to testosterone metabolism and the repeated high-dose exposures required for sufficient enzyme induction.

Clearance mechanisms may be overwhelmed by hepatic or renal toxicity or simply excessively high-dose levels. Altered clearance can affect other toxicokinetic parameters, which can increase systemic blood levels markedly above what would otherwise be expected from a particular dose. Whenever possible, existing toxicokinetic information should be considered when interpreting potential endocrine activity in Tier 1 assays.

Similarly, stress and environmental conditions may change reproductive function irrespective of endocrine activity. For example, Holson et al. (1995) found altered reproductive performance in male rats previously restrained under bright lights, and Stromborg (1986) reported that avian egg production is exquisitely sensitive to food ingestion and adequate nutrition. Of particular concern are potent MoAs, such as cholinesterase inhibition, which can produce specific toxicity that indirectly affects endocrine function at doses well below that which could occur by an endocrine mechanism (Juberg et al., 2013). Care must be taken not to interpret nonspecific responses in the fish short-term reproduction assay (FSTRA; EPA 890.1300), as specific to perturbations in the estrogen androgen, or steroidogenesis pathways, particularly when they are coincident either with signs of overt toxicity or with diagnostic indications of other MoAs apart from endocrine activity (details are explained under specific hypotheses below).

USE OF W_{RELS} IN A WOE INTERPRETATION

A WoE evaluation performed according to the Framework considers the relevance of the Tier 1 ESB data for deciding eight individual hypotheses that the battery was intended to address. Data for a test chemical are evaluated hypothesis by hypothesis. The first step in considering

any particular hypothesis is to evaluate the test chemical responses in the Rank 1 endpoints for that hypothesis. These provide the most direct information about how strongly the hypothesis might be supported for the test substance, or conversely, how limited the support might be. Because the Tier 1 ESB is to be interpreted as a whole, with weaknesses in some assays counterbalanced by the strengths of others, it is important that the evaluation also consider information from Ranks 2 and 3 endpoints and the patterns of response expected for prototype positive and negative controls. However, the response to Rank 1 endpoints should guide the evaluation and interpretation of information from lower ranked endpoints.

Positive responses in Rank 1 are a preliminary indication that the hypothesis is supported. Rank 2 endpoints are then evaluated to determine consistency. Consistent positive responses among Ranks 1 and 2 endpoints can be considered sufficient support for the hypothesis, that is, that the chemical possesses the potential to interact with the specific components of the endocrine system addressed by the hypothesis. Rank 3 endpoints would then be consulted for consistency, and together with the strength of response (W_{RES}) in Ranks 1 and 2 endpoints, temper or strengthen support for the hypothesis. Inconsistent (i.e., negative) responses among Rank 2 endpoints would reduce support for the hypothesis, and depending on the strength of the response in Rank 1 endpoints, could render a conclusion ambiguous in some cases.

A lack of response in Rank 1 endpoints for a particular hypothesis is a preliminary indication that the hypothesis is not supportable. Rank 2 endpoints are then evaluated to determine consistency and completeness. This is particularly important for instances where potential activity might not be explicitly captured by the Rank 1 endpoints. If Rank 2 endpoints are consistent with negative (no response) results on Rank 1 endpoints, a conclusion of the null hypothesis can be rendered with confidence, that is, that the chemical lacks the potential to interact with the specific components of the endocrine system addressed by the hypothesis. Rank 3 endpoints are superfluous for interpretation when findings in both Ranks 1 and 2 endpoints are negative. In this case, positive Rank 3 responses must be considered anomalous or potentially responsive to some other MoA.

The interpretation is more complex if Rank 2 endpoints are inconsistent with negative results in Rank 1 endpoints. In this case, the strength of the response in Rank 2 endpoints becomes even more critical, as does an evaluation of Rank 3 endpoints as well as potential reasons that Rank 1 endpoints may not have responded. It is impossible to discuss how the myriad of inconsistencies between Rank 1 and Rank 2 endpoints might be interpreted, but some overarching themes can be stated. First, Rank 1 endpoints cannot be dismissed for inconsistency with Rank 2. Rank 3 endpoints, in contrast, provide only potentially corroborating information for findings in Ranks 1 and 2 and should not be used alone to evaluate any hypothesis. Situations in which Ranks 2 and 3 are consistent, but inconsistent with Rank 1 endpoints present the greatest challenge, and no general statements can be made. These latter situations must be considered on a case-by-case basis taking into account the strength of the responses, known confounding influences, and

normal biological variation within the various endpoints involved, among other factors.

The Framework does not specifically address hypotheses regarding interactions with the HPT and HPG axis, as no Tier 1 assay is sufficiently specific to those axes. Clearly, the pubertal onset assays as well as the FSTRA and the amphibian metamorphosis assay (AMA; EPA 890.1100; Lutz et al., 2008)) would be expected to respond to agents with such potential, but the specific pattern of results expected is not clear and is subject to variable interpretation. For this reason, the Framework describes a set of hypotheses encompassing apparent thyroid agonism or antagonism (the manifestations of interactions with the HPT axis) as enhanced or reduced thyroid activity. To be sure, there are legitimate reasons to define these hypotheses by different approaches, but irrespective of the approach taken, transparency of the evidence evaluated and decision logic used to make the determination is crucial.

RATIONALE FOR W_{RES} RANKINGS BY HYPOTHESES

The eight hypotheses evaluated by the Tier 1 ESB assays pertain to the potential for a chemical to interact as an (1) agonist with components of estrogen pathways, (2) antagonist with components of estrogen pathways, (3) agonist with components of androgen pathways, (4) antagonist with components of androgen pathways, (5) agonist with components of thyroid pathways, (6) antagonist with components of thyroid pathways, (7) inducer of steroidogenesis enzymes, or (8) inhibitor of steroidogenesis enzymes.

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN AGONIST WITH COMPONENTS OF ESTROGEN PATHWAYS

Estrogenicity is a property that is defined based on a biological response. Estrogens are a class of steroid hormones linked principally with the control of female sex organ responsiveness and of reproduction (Korach et al., 1995). The American Heritage Medical Dictionary defines "estrogenic" as "causing estrus in animals" or "having an action similar to that of an estrogen." This definition guides our evaluation of endpoints for the estrogen agonist hypothesis. Although "estrogens" generally refer to several substances that share a common feminizing activity (Armstrong and Federman, 2005), estrogens influence the growth and functioning of both female and male reproductive tissues (Ososki and Kennelly, 2003). Endpoints from both genders are included in our proposed Rank 1 endpoints for estrogen agonism. The definition of "estrogenic" has been used variably since introduction of in vitro assays capable of probing particular aspects or components of estrogen action, but these are context-dependent (Reel et al., 1996). Here, assays that involve transfer of the test agent via the bloodstream are given more weight than those in which the compound reaches molecular or tissue targets via an artificial fluid or medium. Since endocrine hormones are produced by one tissue and conveyed, usually via the bloodstream, to other organs or tissues where the cellular and physiological effects are produced, and since the purpose of the

Table 1
Estrogen Agonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
FSTRA Vitellogenin: increased in males	ERTA ER agonism	ERBA ER competitive binding affinity
Uterotrophic assay Uterus weight (wet/blotted): increased	FSTRA Secondary sexual characteristics: reduced tubercle score: males Gonad histopathology: males Behavior: males Pubertal female assay Age and weight at vaginal opening: reduced Age at first estrous: reduced Ovary weight: reduced Ovary histopathology: altered Pubertal male assay Testes weight Testes histopathology: atrophy Uterotrophic assay Conversion to estrous supplemental	FSTRA Fecundity Estradiol Testosterone Behavior: females Gonad-somatic index: reduced in males, increased in females Gonad histopathology: follicular atresia Fertilization success Pubertal female assay Growth Estrous cyclicity Pubertal male assay Growth Ventral prostate weight Epididymides histopathology Steroidogenesis assay Estradiol levels

Tier 1 ESB is to determine potential endocrine activity, in vivo assays are given higher priority for this hypothesis. Table 1 lists proposed Ranks 1, 2, and 3 endpoints for the estrogen agonist hypothesis.

Rank 1 Endpoints

In the FSTRA, increased vitellogenesis in males is a specific and sensitive indicator of estrogen agonist potential. During vitellogenesis, the oocyte incorporates vitellogenin (VTG) proteins, lipids, vitamins, and other nutrients into egg yolk proteins (Lubzens et al., 2010.) This culminates in an oocyte competent to undergo fertilization that contains maternal mRNAs, proteins, lipids, carbohydrate, vitamins, and hormones necessary for development of the embryo. The VTG are phosphoglycoproteins primarily synthesized in the liver under the receptor-mediated regulation of 17 β -estradiol and are found in the blood of females of all oviparous vertebrate species. Although several hormones can apparently induce VTG production in fish, many appear to work through or in cooperation with 17 β -estradiol (Lubzens et al., 2010). Vitellogenesis is thus one of the primary reproductive processes under direct estrogenic control. VTG production can also be induced in male fish by administration of exogenous 17 β -estradiol, but in control males, VTG is either not present or is present below the detection limits of most assays (Parks et al., 1999). For this reason, measuring VTG production in males may be a more sensitive indicator of estrogenic activity than an increased production in females. A clear, unambiguous induction of VTG in male fish is a well-established response to estrogen receptor (ER) agonists (Schmid et al., 2002; U.S. EPA, 2007a).

VTG in male fish does not appear to interfere with reproductive success or survival except in rare instances where excessive production leads to kidney failure (Hutchinson et al., 2006; Mills and Chichester, 2008; Caldwell et al., 2012). In vivo assays rather than in vitro or structure-activity relationship models are currently considered the most defensible screening tools for wildlife, in part because of long-standing uncertainties related to xenobiotic metabolism, bioavailability, and toxicokinetics that may not be replicated outside of the intact organism (Ankley et al., 1998). The Panel concluded that increased VTG levels in male fish is a Rank 1 endpoint for evaluating the estrogen agonist potential of a substance.

A statistically and biologically significant increase in wet and blotted uterine weights in the uterotrophic assay (EPA 890.1600; OECD 2007b) in female rats is a sensitive and specific indicator of estrogen agonist potential. Justification for the Rank 1 status of the uterotrophic assay comes both from its biological context and from historical use of the assay. Biologically, estrogens coordinate systemic responses during the ovulatory cycle, including the regulation of the reproductive tract, pituitary, breasts, and other tissues (Fritsch and Murdoch, 1998). Because the principal function of estrogen is to cause cellular proliferation and growth of the tissues of the sex organs and other tissues related to reproduction (Guyton and Hall, 1996a), an in vivo proliferative response is given a higher ranking than other endpoints. Historically, estrogenicity was defined based on the ability to induce uterine growth in immature or ovariectomized rodents (Rozman et al., 2006). Indeed, the crucial evidence establishing the potential estrogenicity of DDT (dichlorodiphenyl-trichloroethane) analogues and other environmental estrogens lies in the observation that their administration

to female animals evokes responses in the uterus or oviduct similar to those observed after administration of classical pharmaceutical estrogens, such as 17β -estradiol (Bulger and Kupfer, 1985). The uterotrophic effects of estrogen agonists are completely blocked by a pure ER antagonist (Wade et al., 1993). Results of OECD validation studies for the uterotrophic assay have been published, establishing its ability to discriminate true positives from negatives with respect to potential for estrogenic activity in vivo (Odum et al., 1997; Kanno et al., 2001, 2003; Owens and Ashby 2002; Owens et al., 2003). The main advantage of the uterotrophic assay is the determination of an overall biological effect that can account for interactions between cells and between different components of the endocrine system (Wang et al., 2012) as well as absorption, distribution, metabolism, and excretion (ADME).

For these reasons, the uterine weight (wet and blotted) endpoint in the rat uterotrophic assay is assigned a Rank 1 relevance weighting. Despite the strength of information provided by the uterotrophic assay for deciding the estrogen agonist hypothesis, it should be noted that uterotrophic responses are possible from other mechanisms, such as the anabolic uterotrophic action by androgens, which is blocked by antiandrogens but not by antiestrogens (Schmidt and Katzenellenbogen, 1979; Beri et al., 1998; Wang et al., 2012).

Rank 2 Endpoints

The agonist arm of the ERTA assay (EPA 890.1300) provides a measure of activation by a test substance that is highly specific for the alpha (α) form of the human ER (hER α). Responses in the assay are mathematically comparable to the level of activation produced by a prototypical full ER agonist, such as 17β -estradiol, and so can provide information about strength of response. The assay can also detect the ability of a test substance to inhibit the activity of an agonist, and so can distinguish between agonists, partial agonists, and antagonists. Although the ERTA assay is highly specific and sensitive for hER α , it does not provide a measure of response through the beta form, hER β . It is subject to the general limitations of in vitro assays (Rozman et al., 2006), and is therefore less informative for the estrogen agonist hypothesis than Rank 1 endpoints.

Several endpoints in male fish (FSTRA) are also considered interpretable for estrogen agonist potential, including decreased tubercle scores in male fish, specific alterations in male gonad histopathology (e.g., degenerate spermatozoa and proliferation of Sertoli cells) and altered male behavior, specifically the expectation that aggressive male behavior and nest guarding activities be significantly reduced when exposed to estrogen agonists (Shappell et al., 2010). These characteristics are relevant, reliable, and can be indicative of estrogenic activity (OECD, 2007a). However, these endpoints can also be confounded by other hormonal activities or systemic toxicity and thus, are less specific than VTG production (Clearwater and Pankhurst, 1997; Pankhurst and Van Der Kraak, 2000). Furthermore, these endpoints are largely subjective, requiring expert judgments that render them more susceptible to methodological variability and less useful for measuring strength of response. For these reasons, the Panel

assigned these endpoints Rank 2 status for evaluating estrogen agonist potential.

Several endpoints measured in the pubertal female assay are proposed as Rank 2 for evaluation of estrogen agonist activity because of their direct dependence on 17β -estradiol. These include biologically relevant age advancement (≥ 2 days younger) and decreased body weight at VO, biologically significantly decreased ovarian weights and altered ovarian histopathology (Marty et al., 1999; Kim et al., 2002; Stoker and Zorrilla, 2010). However, these endpoints may not respond consistently for the same test chemical (Marty et al., 1999; Stoker and Zorrilla, 2010) may also respond to chemicals with other hormonal activity (Kim et al., 2002), and the lack of clear negative results for a set of test compounds known to lack hormonal activity decreases confidence in the interpretations that can be placed in positive results. Some endpoints may respond generally to stress. Thus, these endpoints may not be as clearly interpretable or as useful for assessing the presence of estrogenic activity or measuring the strength of estrogen-like responses and are therefore proposed as Rank 2 rather than Rank 1 endpoints for evaluation of estrogen agonist potential.

In the pubertal male assay, testes weight and testes atrophy (demonstrated by testes histopathology) are proposed as Rank 2 for evaluation of estrogen agonist activity. Although estrogens—that is, 17β -estradiol—interfere with testes development by causing testicular atrophy, this endpoint is not specific for estrogens; other MoAs, such as oxidative stress, can also cause testicular atrophy. Histological evaluation may help in interpreting a response, but, as with the female pubertal assay, consistency of response to estrogens in the male pubertal assay is not expected, even for the same compound (Marty et al., 1999; Ashby and Lefevre, 2000; Stoker and Zorrilla, 2010).

OECD's validation of the uterotrophic assay considered several optional endpoints, but none were found to be more sensitive or specific than an increase in uterine weight for evaluation of estrogen agonist potential. Although conversion to estrus is an endpoint specifically under estrogenic control, its assessment is not as quantitative or objective as measurement of uterine weight, and therefore, may be more susceptible to methodological confounding than the Rank 1 endpoint.

Rank 3 Endpoints

The ERBA is an in vitro assay that determines the ability of the test chemical to bind with rat uterine cytosolic ER and is very sensitive and specific for this purpose (Kuiper et al., 1997). However, it is considered a Rank 3 endpoint, rather than Rank 2, because it does not allow the interpretation of agonist versus antagonist properties (Black et al., 1981; Kuiper et al., 1997). For estrogen agonist potential, both binding affinity and intrinsic activity (efficacy) at the receptor are required. Due to the high degree of conservation across mammals, compounds that bind to rat ER are assumed to be capable of binding to human ER (EPA 890.1250). ER binding alone does not consider the complex biological systems of a whole organism in determining a positive response. Metabolism of the test article in vivo and ER interaction with other cellular factors are not evaluated in the assay. Therefore, the assay cannot

determine if binding will result in biological activity. A range of compounds can affect the results of the assay by altering buffer pH, denaturing the ER, or disrupting the ER binding kinetics, none of which are relevant to potential endocrine activity. Thus, several conditions may confound interpretation of ERBA data and require secondary diagnostic analyses to confirm competitive inhibition (Laws et al., 2006). Although there can be technical difficulties with interpretation of the assay and corroboration from other assays is required to determine the activity of a substance in this assay, it is considered a strong corroborative assay for substances that exhibit estrogen agonist (or antagonist) potential in Ranks 1 and 2 endpoints.

Several endpoints in the FSTRA may respond to estrogenic substances, albeit with lower specificity and sensitivity compared to Ranks 1 and 2 endpoints. The FSTRA endpoints proposed as Rank 3 include gonad-somatic index (GSI), fecundity, estradiol and testosterone levels, behavior in females, some gonadal histopathology findings (i.e., follicular atresia), and fertilization success. The mechanisms underlying these responses are not well characterized and may not be specific to estrogens. Reduction in sex steroid concentrations, GSI, reproduction, and altered gonadal histopathology (particularly when seen as isolated or nearly isolated "positive" findings) can be the result of general systemic toxicity unrelated to estrogenic activity or to perturbation of the HPG axis, androgenic, and/or steroidogenesis pathways (Foo and Lam, 1993; Clearwater and Pankhurst, 1997; Haddy and Pankhurst, 1999; Lethimonier et al., 2000; Pankhurst and Van Der Kraak, 2000; Wu et al., 2003; Aluru and Vijayan, 2009; Milla et al., 2009). Fecundity is particularly sensitive to systemic toxicity, decreased feeding or stress, and when seen in isolation, is a poor predictor of potential endocrine activity. Therefore, the Panel considered the interpretation of responses in these endpoints to depend on the response of other endpoints, and thus, to be of corroborative use only.

In the pubertal female assay, estrogen agonists promote growth and estrous cyclicity. These endpoints, however, were considered to be insufficiently specific to be interpretable without the context of other assays, may require high doses of potent estrogens to elicit responses, and may be susceptible to nonendocrine MoAs (Marty et al., 1999; Stoker and Zorrilla, 2010). They were therefore categorized as Rank 3 for the estrogen agonist hypothesis.

Similarly, growth, ventral prostate weight, and histopathology of the epididymides in the pubertal male assay respond to estrogen agonists, but the responses lack specificity (Stoker and Zorrilla, 2010) and may be susceptible to nonendocrine MoAs. These were therefore also given Rank 3 weightings for the estrogen agonist hypothesis.

It is important to appreciate that most of the endpoints measured in the male and female pubertal onset assays are considered Rank 3 rather than Rank 2. The pubertal assays are multimodal assays that rely on apical endpoints such as tissue and organ weights that will respond to more than one endocrine hormone. Although this aspect is a strength in that the assays may detect endocrine signals that are not detected by other assays, this also makes it difficult to determine the MoA because the endpoints are not specifically modulated by a single hormone. Fur-

thermore, responses in these endpoints may reflect secondary, nonspecific effects related to systemic toxicity, stress, or other factors unrelated to hormonal signals. It is for these reasons that several endpoints in the pubertal assays are given Rank 3. Clearly, interpretation of these endpoints depends on more specific responses from other assays.

Compounds that increase estradiol levels in the steroidogenesis assay may produce responses in other assays that suggest a potential for estrogen agonism due to an increased level of the endogenous hormone rather than a direct agonist effect of the substance. Because circulating estradiol levels are controlled by a variety of factors, results in this assay can only provide corroboration and perhaps clarification for Ranks 1 and 2 endpoint responses.

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN ANTAGONIST WITH COMPONENTS OF ESTROGEN PATHWAYS

The biological properties of estrogens are summarized under the estrogen agonist hypothesis. Antiestrogens were classically defined as substances that inhibit the growth-promoting effect of estradiol on target tissues (Black et al., 1981). Antagonist activity can occur by direct interaction with ERs or by indirect mechanisms such as reduction of circulating estradiol levels or by enhancement of the activity of a hormone or other regulatory molecule whose biological actions oppose those of endogenous estrogens (MacGregor and Jordan, 1998). As with agonists, receptor-mediated antagonist potential is directly related to receptor affinity. In contrast with estrogen agonists, estrogen antagonist potential is inversely related to the intrinsic efficacy of the substance to activate the ER and produce cellular responses (Black et al., 1981). Such substances may be "pure" receptor antagonists, which oppose the activity of 17 β -estradiol at the ER but produce no response alone, or may be "partial" or "inverse" agonists, both of which can reduce the "estrogenic tone," that is, the background hormonal or constitutive ER activity. The extent to which estrogenic tone is reduced depends on the properties of the substance and the conditions of the particular assay used to measure the activity (reviewed by Negus, 2006). Regardless of these nuances, chemicals with sufficient ER affinity but with intrinsic efficacy less than that of 17 β -estradiol have the potential to reduce estrogenic activity—that is, to act as estrogen antagonists—under some conditions. Table 2 lists proposed Ranks 1, 2, and 3 endpoints for the estrogen antagonist hypothesis.

Rank 1 Endpoints

Currently, the Panel did not consider any of the current Tier 1 endpoints sufficiently specific and reliable to assign them a Rank 1 W_{REL} weighting for potential antiestrogenic activity. If fully validated, the antiestrogenic mode of the uterotrophic assay would qualify as a Rank 1 endpoint for potential antiestrogenic effects. The antiestrogenic procedure replaces physiological estrogen in ovariectomized female rats by administration of a uterotrophic dose of 17 β -estradiol (estrogen control). Treatment groups then receive the test chemical (treated). Attenuation of the uterotrophic effect of the exogenous 17 β -estradiol by the

Table 2
Estrogen Antagonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
	ERBA ER competitive binding affinity	Aromatase assay Aromatase inhibition
	Pubertal female assay Age and weight at vaginal opening: increased Age at first estrus: increased	Steroidogenesis assay Estradiol levels
	FSTRA Vitellogenin: reduced in females Gonad histopathology: in females	Pubertal female assay Estrous cyclicity Ovary histopathology (atrophy) Ovary weight: reduced, with atrophy
		FSTRA Fecundity Estradiol Testosterone Gonad-somatic index Behavior Fertilization success

test chemical (treated vs. estrogen control) indicates potential antiestrogenic activity.

Rank 2 Endpoints

The ERBA is an *in vitro* assay that determines the ability of the test chemical to bind with rat uterine cytosolic ER and is very sensitive and specific for this purpose (Kuiper et al., 1997). The ERBA specifically determines the specific binding affinity of a chemical for the ER under the conditions of the assay, but does not allow the interpretation of agonist versus antagonist properties (Black et al., 1981; Kuiper et al., 1997). Because intrinsic activity is not required for antiestrogenic potential, the determination of specific receptor binding affinity is more informative for estrogen antagonism versus agonism, and therefore, the ERBA is given a higher ranking for the estrogen antagonist versus estrogen agonist hypothesis. Due to the high degree of conservation across mammals, compounds that bind to rat ER are assumed to be capable of binding to human ER (EPA 890.1250). ER binding is categorized as a Rank 2 endpoint rather than Rank 1 because it does not consider the complex biological systems of a whole organism in determining a positive response. Metabolism of the compound *in vivo* and ER interaction with other cellular factors are not considered in the result. Therefore, the assay cannot determine if binding will result in biological activity. Additionally, without corroboration from other assays, it is not possible to determine if the compound is capable of an agonistic versus antagonistic activity.

In the pubertal female assay, substances with potential antiestrogenic activity increase the age and weight at VO and the age at which the first estrus occurs (≥ 2 days older). Limitations of these endpoints are discussed under Rank 2 endpoints for the estrogen agonist hypothesis (Section 1.2). Although these endpoints respond to a generalized decrease in estrogenic tone, this can occur via many different mechanisms, including antagonism at the

ER and inhibition of aromatase or gonadotropin, leading to decreased estradiol levels (Marty et al., 1999; Ashby et al., 2002; Stoker and Zorrilla, 2010). Responses may require relatively potent compounds or high doses, the latter of which increases the chance of generating confounding results due to systemic toxicity.

Chemicals with potential antiestrogenic activity can affect the FSTRA by decreasing female VTG (U.S. EPA 2007a). Antiestrogenic reduction of VTG in female fish is not as specific, reliable, or robust as the estrogen agonist induced VTG increase in male fish. VTG reduction in female fish can also occur from stress and conceivably from hepatotoxicity and crosstalk with the aryl hydrocarbon receptor (Anderson et al., 1996a, 1996b; Lethimonier et al., 2000; Celander, 2011; Bugel et al., 2013). Thus, reduced VTG in female fish is placed in Rank 2 for the antiestrogenic hypothesis. Gonad histopathology was also given Rank 2 status because as the primary reproductive organ, the gonad produces physiologic responses to endocrine-active substances (Ankley et al., 2001; U.S. EPA, 2006). While there is little data on the expected profile of an antiestrogen in the FSTRA (U.S. EPA, 2007a), it is assumed that decreases in estrogenic activity that result in decreased VTG levels could also result in observable alterations to ovarian histopathology in female fish. Thus, in light of the scarcity of sensitive endpoints for indicating antiestrogenic activity in the FSTRA, decreases in female VTG and alterations in ovarian histopathology that are consistent with decreases in estrogenic activity are designated as Rank 2 endpoints.

Rank 3 Endpoints

Inhibition of aromatase (aromatase assay; EPA 890.1200) prevents conversion of 19C-androgens to 18C-estrogens and hence, conversion of testosterone to estradiol, resulting in lower circulating 17 β -estradiol levels (Marty et al., 2001b). If significant and persistent, lower circulating levels of 17 β -estradiol might reduce estrogenic tone. On the other hand, since conversion of androgens to estrogens via aromatase is not the primary source of estrogens, especially in humans, and because estrogen sulfates may be converted to free estrogens in compensation for reduced estrogen levels, a modest decrease in aromatase activity would have little if any effect on steady-state estrogen levels. Thus, inhibition of aromatase may suggest that a chemical has the potential to exhibit antiestrogenic effects, but the response in this assay could only corroborate antiestrogenic activity in Ranks 1 and 2 assays.

Similarly, compounds that reduce estradiol levels in the steroidogenesis assay may produce responses in other assays that suggest a potential for estrogen antagonism due to reduced levels of the endogenous hormone rather than a direct receptor antagonist effect of the substance. Because circulating estradiol levels are controlled by a variety of factors, however, results in this assay can only provide corroboration and perhaps clarification for Ranks 1 and 2 endpoint responses.

Antiestrogens may alter estrous cyclicity in the pubertal female assay, and may induce ovarian atrophy as evidenced by a concordance of atrophic changes in histopathology and reduced organ weight. Although

these endpoints are not dispositive as a unique response, it may be used to corroborate more specific responses from other assays, and is therefore assigned Rank 3 priority.

Several endpoints in the FSTRA (fecundity, estradiol, and testosterone levels, GSI, behavior, and fertilization success) may respond to potential antiestrogens. However, as noted under the estrogen agonist hypothesis, these endpoints are difficult to interpret because of their responsiveness to different hormonal activities as well as their significant potential to be confounded by general systemic toxicity that also affects endocrine sensitive tissues. For these reasons, Rank 3 priority is appropriate for these as corroborative endpoints only for antiestrogenic potential.

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN AGONIST WITH COMPONENTS OF ANDROGEN PATHWAYS

Androgens are defined as any steroid hormone that controls the development and maintenance of male sexual organs and male secondary sex characteristics (American Heritage Medical Dictionary, 2007). Like estrogens, androgens are present in both males and females. In females, the production of androgens promotes the development of secondary sex characteristics that occur during puberty (Hiipakka and Liao, 1995; Guyton and Hall, 1996b). Typically, androgens are released into the blood stream where they bind tightly to a serum glycoprotein of hepatic origin called SHBG in humans or, weakly, to albumin in humans and rats (Winters, 1998; Kicman, 2010; Hammond, 2011). The bound hormone is delivered to and enters target cells where it binds to the androgen receptor (AR) and stimulates the production of androgen-dependent proteins (Winters, 1998; Kicman, 2010). The majority of steroid-induced androgenic effects result from an increased rate of protein formation in these target cells. For example, testosterone in prostate tissue induces the DNA–RNA transcription process. After several days, a simultaneous increase in the number of prostatic cells and quantity of DNA can be identified in the gland (Guyton and Hall, 1996b). The effects of androgens on target tissues are not always inductive. Androgens inhibit the pituitary alpha-subunit gene expression and gonadotropin-releasing hormone released by the hypothalamus (Winters, 1998). Table 3 lists proposed Ranks 1, 2, and 3 endpoints for the androgen agonist hypothesis.

Rank 1 Endpoints

The androgenic mode of the Hershberger assay (EPA 890.1400) measures responses to the test article in castrated male rats. Several endpoints are considered Rank 1 because this *in vivo* assay measures specific tissue responses that are hallmarks of androgenic action and incorporates the ADME of the chemical. In the castrated male rat model, a concordant increase in the absolute weights of five androgen-dependent tissues—Cowper's gland, seminal vesicles, levator ani-bulbocavernosus muscle (LABC), glans penis, and ventral prostate—is considered a clear indication of potential androgenic activity (OECD, 2009a). Although androgens are known to increase the

weight of these tissues, tissue growth can be altered by MoAs other than androgen agonism. Therefore, a significant growth response from multiple androgen-dependent tissues is necessary to suggest an androgenic response (OECD, 2009a). A statistically significant increase in organ weight for two or more of the assay tissues is required for a positive result according to the EPA test guideline (U.S. EPA, 2011b). Confidence that the chemical has androgen agonist potential increases as the number of significantly affected organs increases, and decreases as fewer endpoints respond. The Panel therefore assigned a concordant response in all five endpoints Rank 1 for the androgen agonist hypothesis.

In the FSTRA, the appearance of male secondary sex characteristics in female fish can provide a clear indication of androgenic potential (Ankley et al., 2003; U.S. EPA, 2007a; Ankley and Gray, 2013). As in mammals, the secondary sex characteristics in fish are controlled by sex steroids and, therefore, can be altered by exposure to endocrine-active compounds. In female fish of some species, androgen agonists induce the masculinization of secondary sex characteristics (U.S. EPA, 2007a). Exposure to an androgen agonist results in abnormal sexual differentiation and function in female fathead minnows, including the development of nuptial breeding tubercles, which normally are expressed only in male fathead minnows (Ankley and Villeneuve, 2006). In fathead minnows, the development of nuptial tubercles in females is considered diagnostic of androgen exposure (OECD, 2009b). Therefore, the development of tubercles in female fathead minnows was assigned a Rank 1 endpoint since it is reproducible and biologically relevant (U.S. EPA, 2006).

Rank 2 Endpoints

The ARBA measures competitive binding affinity of the test article for rat prostatic cytosol ARs *in vitro*. *In vivo*, this molecular interaction initiates the cellular response to endogenous androgens. Due to the high degree of conservation across mammals, compounds that bind to rat AR are assumed to be capable of binding to AR in humans (U.S. EPA, 2011c). Although receptor binding potential can indicate a potential interaction with androgenic pathways, binding alone cannot distinguish between potential agonist versus antagonist properties. Without corroboration from other assays, it is not possible to determine if the compound has potential for agonistic versus antagonistic responses. Furthermore, the assay cannot determine if binding will result in a biologically significant endocrine effect because ADME of the compound *in vivo*, and AR interactions with other cellular factors are not necessary for the *in vitro* result. The Panel would have assigned the ARBA to Rank 3 for the androgen agonist hypothesis, similar to the ranking for the ERBA for estrogen agonism, if a validated AR transactivation assay were a component of the Tier 1 battery. However, lacking another assay that allows this determination at the molecular level, the Panel assigned the ARBA Rank 2 status until it can be replaced by a fully validated *in vitro* assay more specific for agonist potential.

As stated above, significant growth responses in multiple androgen-dependent tissues (Cowper's gland, seminal vesicles, LABC, glans penis, and ventral prostate) in

Table 3
Androgen Agonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
Hershberger assay Concordance of 5 endpoints: Cowper's gland weight Seminal vesicle weight LABC weights Glans penis weight Ventral prostate weight	ARBA AR competitive binding affinity	Aromatase assay Aromatase inhibition
FSTRA Secondary sexual characteristics: tubercles in females	Pubertal male assay Age and weight at preputial separation: if accelerated Seminal vesicle + coagulating gland weight (wet/blotted) Ventral prostate weight Dorsolateral prostate weight Levator ani-bulbocavernosus muscle complex weight Epididymis weight Testes weight Testes histopathology Epididymides histopathology	Steroidogenesis assay Testosterone levels
	FSTRA Vitellogenin: reduced in females Gonad histopathology	Pubertal male assay Growth Testosterone levels
	Hershberger assay Concordance of two or more endpoints (see Rank 1)	Pubertal female assay Growth Age and weight at vaginal opening Uterus weight Ovary weight Adrenals weight Uterus histopathology Ovary histopathology
		FSTRA Fecundity Testosterone and estradiol levels Gonad-somatic index Behavior Fertilization success
		Hershberger assay Only one of five endpoints respond (see Rank 1)

the androgenic mode of the Hershberger assay suggests androgenic potential (OECD, 2009a). The interpretation is most clear when several tissues respond concordantly. The EPA requires a statistically significant response from two or more of the tissues to interpret a positive response (U.S. EPA, 2011b), and therefore, the Panel assigned concordance of only two of the five endpoints Rank 2 for the androgen agonist hypothesis.

The pubertal male assay is used to identify potential endocrine activity underlying pubertal development and thyroid function in male rats. It consists of collections of *in vivo* endpoints that are interpreted according to expected patterns of response that together indicate a specific hormonal potential, rather than as individual endpoints. The response profile from the endpoints in the pubertal male assay reflects a potential for specific endocrine modalities (e.g., androgen agonism or antagonism, alteration of steroidogenesis). However, a chemical may not alter all endpoints indicative of a particular hormonal mechanism or may alter some endpoints but not others, depending on the dose. Due to variation of response profiles for some chemicals, U.S. EPA (2009) does not require consistency among endpoints (even redundant ones) to make interpretations. However, inconsistency in the response profile for a given chemical decreases confidence in the interpretation. Nonetheless, certain responses in the pubertal male assay are sufficiently interpretable (described below) to warrant placement in Rank 2.

In the pubertal male assay, age and weight at PPS are considered markers for the onset of puberty. Because 5 α -DHT initiates PPS, the endpoint is androgen-

dependent and specific for identifying endocrine activity (Stoker et al., 2000; U.S. EPA, 2011b). A decrease in age at PPS has been observed with the administration of exogenous methyl testosterone to male rats, indicating that acceleration of PPS is a sensitive and specific marker of androgen agonism (Stoker et al., 2000). Androgens are necessary for the development and maintenance of the organs in the male reproductive tract and hence, the weights of several glands and muscles within this tract are considered markers of androgenic action. These include seminal vesicle and coagulating gland (wet and blotted), ventral prostate, dorsolateral prostate, LABC muscle complex, epididymis, and testes. Exposure to methyl testosterone in the pubertal male rat usually causes a significant increase in reproductive organ weights (Stoker et al., 2000; Marty et al., 2001b). However, in the validation study conducted by U.S. EPA (2007b), administration of 80 mg/kg methyl testosterone significantly increased the weight of the ventral prostate and seminal vesicle and decreased the weight of the testes and epididymides. The decrease in weight of these organs is hypothesized to be a result of downregulation of gonadotropins and, subsequently, luteinizing hormone (LH) in the hypothalamus (Stoker et al., 2000). According to this hypothesis, increased male reproductive organ weight in the pubertal male rat is specific to androgen agonists but a decrease in testes and epididymides weight may also indicate androgen agonist potential. On the other hand, aromatization of methyl testosterone to estradiol has not been ruled out, as aromatase has been found to be significant in testes and epididymides in Rhesus monkey

(Pereyra-Martinez et al., 2001; Shayu and Rao, 2006). The dichotomy in response confounds interpretation of the results, suggesting a Rank 2 classification.

In a prevalidation study for the male pubertal assay (Rocca and Pepperl, 2000), rats exposed to methyl testosterone showed significant changes in histopathology of the testes and epididymides. The testes displayed hypospermatogenesis and interstitial cellular atrophy. The epididymides displayed hypospermia secondary to a decreased production in the testes and degenerative germ cells due to release of degenerated cells from the testes. These findings are consistent with exposure to methyl testosterone and indicate androgen agonist potential. Thyroid antagonists also produce histological changes in the testes and epididymides, which confounds interpretation of the results (Rocca and Pepperl, 2000).

In the FSTRA, the Panel assigned reduced VTG in females Rank 2 status. Exposure to androgenic chemicals decreases the production of endogenous androgens, which are converted to estradiol (U.S. EPA, 2006). In female fathead minnows, VTG is produced in response to estrogenic compounds, including estradiol (Ankley et al., 2001). Therefore, the lower levels of estradiol result in decreased levels of VTG, as has been observed in response to 17 β -trenbolone exposure (Ankley and Gray, 2013). While decreased VTG can indicate a specific endocrine activity, there are confounding factors that complicate interpretation. Production of VTG in female fish can also be decreased by hepatotoxicity, stress, and other nonendocrine modes of toxicity (Anderson et al., 1996a, 1996b; Lethimonier et al., 2000; U.S. EPA, 2006; OECD, 2009b; Celander, 2011; Bugel et al., 2013). Gonad histopathology was also given Rank 2 status because as the primary reproductive organ, the gonad produces physiologic responses to endocrine-active substances (Ankley et al., 2001; U.S. EPA, 2006). In response to androgenic chemicals, observations of hyperproduction of sperm in male fish and a decrease in yolk deposition in female fish have been observed (U.S. EPA, 2007a). These alterations in gonadal histopathology were classified as Rank 2 endpoints because while these changes may result from endocrine exposure, they are not necessarily specific to these substances. This confounds interpretation of the results. Additionally, despite standardization of histopathological primary diagnoses, they are still subject to individual interpretation and may reflect observation bias (U.S. EPA, 2006).

Rank 3 Endpoints

As stated above, significant growth responses in multiple androgen-dependent tissues (Cowper's gland, seminal vesicles, LABC, glans penis, and ventral prostate) in the androgenic mode of the Hershberger assay suggests androgenic potential (OECD, 2009a). The interpretation is most clear when several tissues respond concordantly, but is much less clear if only one tissue shows a growth response. Because the EPA requires a statistically significant response from two or more of the tissues to interpret a positive response, a single responding tissue is not considered sufficient evidence of androgenic activity (U.S. EPA, 2011b). A single responding tissue was classified as a Rank 3 endpoint because it produces only qualitative

evidence of androgenic potential. It should be evaluated with additional endpoints to determine the relevance of the response.

Inhibition of aromatase in the aromatase assay is reflected by decreased estradiol and increased testosterone levels. The aromatase assay utilizes human placental recombinant microsomal aromatase to measure aromatase enzyme activity *in vitro*. Aromatase is a cytochrome P450 (CYP) complex that metabolizes androgens to estrogens. Therefore, alteration of aromatase activity levels affects the amount of estrogen present in the tissues and, subsequently, alters endogenous hormone levels and activity. Androgen agonists downregulate aromatase gene expression and subsequently inhibit their activity (Lanzino et al., 2013). While a significant decrease in aromatase activity is characteristic of androgen agonism, it is not definitive for this hypothesis. Other mechanisms of action, including cytotoxicity, will produce an apparent decrease in aromatase activity (Battelle, 2005). Inhibition of aromatase is classified as a Rank 3 assay because it neither determines the mechanism of action nor predicts effects at the protein level postprocessing. As a result, additional lines of evidence are needed to corroborate androgen agonism.

In the steroidogenesis assay, testosterone levels were assigned Rank 3 status for the androgen agonist hypothesis. This *in vitro* assay detects the synthesis of steroid hormones by measurement of their concentrations in a human adrenocortical carcinoma cell line, H295R. As with all *in vitro* assays, the steroidogenesis assay has limited ability for metabolism and cannot predict activity at the organism level. While the steroidogenesis assay can corroborate the results of other assays, it cannot establish whether the chemical alters activity at the AR. Androgen agonists, such as trenbolone acetate, significantly decrease testosterone concentrations in H295R cells (Gracia et al., 2007). The mechanism for this is unclear, but may be due to downregulation of 17 β -hydroxysteroid dehydrogenase, an enzyme responsible for the conversion of androstenedione to testosterone (Hilscherova et al., 2004).

In the pubertal male assay, growth and circulating levels of testosterone were assigned Rank 3 status because although androgen agonists are known to significantly decrease body weight and serum testosterone and DHT levels, there are several confounding factors for these endpoints. Androgen agonists cause feedback inhibition of gonadotropin release and a subsequent decrease in testosterone production by the Leydig cells (O'Connor et al., 2000). This is especially true for growth, which is affected by a variety of factors including overt toxicity, food intake, and onset of puberty (Laws et al., 2007; Marty, 2013). Without corroboration from other assays, the effects cannot confidently be attributed to endocrine agonism.

Numerous endpoints in the pubertal female assay were assigned Rank 3 status, including growth, age, and weight at VO, uterine, ovarian, and adrenal gland weights, and uterine and ovarian histopathology. Similar to the pubertal male assay, the pubertal female assay is used to identify potential endocrine activity underlying pubertal development and thyroid function. It is a collection of *in vivo* endpoints that are interpreted according to expected response profiles that together indicate a potential

for endocrine activity (e.g., estrogen agonist, antiestrogen). The pubertal female assay is not often utilized for the assessment of androgens and very little information is available on the subject. Additionally, the U.S. EPA did not utilize an androgen agonist during the pubertal female validation studies, making interpretation of the assay more difficult for this group of chemicals (U.S. EPA, 2007c). Chemical-specific variations in response also confound the analysis. The endpoints are classified as Rank 3 because they cannot indicate androgen agonism without corroboration from other assays.

Body weight is significantly increased in female pubertal rats exposed to androgen agonists (Kim et al., 2002; Clark et al., 2003). Body weight is classified as a Rank 3 endpoint because it can be affected by food intake and overt toxicity, confounding interpretation. At low concentrations (near physiological levels), exposures to aromatizable androgen agonists decrease the age and weight at VO. Nonaromatizable androgen agonists have no effect on this endpoint. Because serum estradiol concentrations are not increased after exposure to aromatizable androgens, it is hypothesized that early VO is the result of local aromatization to estrogen and direct action on the vaginal epithelium (Goldman et al., 2000). Higher concentrations of aromatizable androgens have the opposite effect. Administration of 1 mg/kg/day testosterone delayed VO and increased body weight at VO (Kim et al., 2002). This endpoint was classified as Rank 3 because the response is not dose-dependent and can be affected by changes in growth. Additionally, other mechanisms (e.g., estrogen agonists) produce similar results, that is, early VO, and may confound interpretation of the response (Goldman et al., 2000).

Both uterine and ovarian weights were reduced by administration of testosterone at 1 mg/kg/day (Kim et al., 2002). However, the effects on the uterus appear to be dose dependent since uterine weight was significantly increased after administration of levels above 20 mg/kg/day testosterone due to uterine stromal cell proliferation (O'Connor et al., 2000; Wason et al., 2003), an effect that may also be mediated via ARs (Schmidt and Katzenellenbogen, 1979; Beri et al., 1998; Wang et al., 2012). Because the pubertal female assay uses intact rats, cyclic changes in uterine weight also affect the variability of this endpoint, confounding interpretation. Histopathology results, particularly regarding stromal proliferation, may be necessary to draw conclusions. Adrenal weight to brain weight ratio is significantly increased by exposure to the androgen agonist 17 α -methyltestosterone at high-dose levels (Wason et al., 2003), however, no change in adrenal weight is noted at lower exposure concentrations (Kim et al., 2002). Interpretation of this endpoint is confounded by the dependence on high doses, which could also conceivably produce a stress response. Exposure to androgen agonists results in atrophy of the ovary and inactivation of the interstitial glands. The number of follicles and corpora lutea may also be reduced. In contrast, the uterus displays hyperplasia and hypertrophy of the epithelium (Wason et al., 2003).

In the FSTRA, fecundity, testosterone and estradiol levels, GSI, behavior, and fertilization success were classified as a Rank 3 because they are relevant, but not specific to

androgen agonism. These endpoints are not considered indicative of endocrine action unless Ranks 1 and 2 endpoints respond. The number of eggs spawned is mediated by the endocrine system. Changes in hormonal equilibrium, gonadal development, and vitellogenesis will also affect fecundity (U.S. EPA, 2006). If higher ranked endpoints are not positive for endocrine MoAs, the effects on fecundity cannot be definitively identified as evidence of androgen agonism (OECD, 2004a). Androgen agonists may increase or decrease or have no effect on testosterone and estradiol concentrations in fish (U.S. EPA, 2007a). Circulating levels of sex steroids can vary seasonally or even daily, confounding interpretation of the results. Due to the variety of causes, changes in the concentration of sex steroids can support Ranks 1 and 2 endpoints but have little value on their own (U.S. EPA, 2006). The GSI response to an androgen agonist, however, was less predictable with either no effect on GSI or increased or decreased GSI observed among male and female fish in response to androgen agonists (U.S. EPA, 2007a). A primary assumption for this endpoint is that the relationship between gonad and body weight is constant. This does not hold true during spawning as female GSI can vary as much as 45% over a 2-day period, confounding interpretation of the results (OECD, 2004a). Territorial aggressiveness and changes in reproductive behavior have been observed in female fathead minnows exposed to androgens (Ankley et al., 2001; OECD, 2009b). However, changes in behavior have a variety of causes including exposure to non-endocrine-active chemicals and systemic toxicity. Additionally, behavior is not routinely quantitated or systematically observed and incorporates subjectivity in interpretation of the endpoint (OECD, 2004a). Based on the nonspecific nature of this endpoint, it was classified as Rank 3 and should not be utilized as the only indicator of potential endocrine pathway interaction. Decreased VTG and disturbances in the normal levels of sex hormones decrease fertilization success. Because changes in sex hormones and decreases in VTG may be caused by nonendocrine MoAs, fertilization success is not specific for endocrine-active compounds. Additionally, fertilization measures the combined effects from both genders, making it difficult to assess gender-specific sensitivities (OECD, 2004a).

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN ANTAGONIST WITH COMPONENTS OF ANDROGEN PATHWAYS

The U.S. National Library of Medicine (2013) defines androgen antagonists as compounds that inhibit or antagonize the biosynthesis or actions of androgens. Direct acting androgen antagonists bind to the AR in target tissues and block androgen-induced receptor activation and transcription of androgen-dependent genes (Winters, 1998). Indirect acting androgen antagonists reduce intracellular steroid levels by mechanisms other than binding to the receptor, including decreasing the rate of steroid synthesis, reducing the plasma-free fraction, and increasing the rate of androgen elimination (OECD, 2004a). Table 4 lists proposed Ranks 1, 2, and 3 endpoints for the androgen antagonist hypothesis.

Table 4
Androgen Antagonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
Hershberger assay Concordance of 5 endpoints: Cowper's gland weight Seminal vesicle weight LABC weights Glans penis weight Ventral prostate weight	ARBA AR competitive binding affinity Pubertal male assay Age and weight at preputial separation: if delayed (age and weight increased) Seminal vesicle + coagulating gland weight (wet/blotted) Ventral prostate weight Dorsolateral prostate weight Levator ani-bulbocavernosus muscle complex weight Epididymis weight Testes weight Testes histopathology Epididymides histopathology FSTRA Vitellogenin: increase in females Secondary sexual characteristics: reduced in males Gonad histopathology Hershberger assay Concordance of two or more endpoints (see Rank 1)	Steroidogenesis assay Testosterone levels Pubertal male assay Testosterone levels FSTRA Fecundity Estradiol Testosterone Gonad-somatic index Behavior Fertilization success Hershberger assay Only one of five endpoints respond (see Rank 1)

Rank 1 Endpoints

The antiandrogenic mode of the Hershberger assay measures responses to the test article following replenishment of androgen to castrated male rats by administration of testosterone propionate. Several organ weight endpoints measured in the antiandrogenic mode of the Hershberger assay—Cowper's gland, seminal vesicles, LABC, glans penis, and ventral prostate—are given Rank 1 status for the androgen antagonist hypothesis when a concordant response is produced. The reasons are the same as those explained for the androgen agonist hypothesis. Specifically, the Hershberger is an *in vivo* assay that measures responses in tissues that define the subject hormonal activity, and it is consistent and reproducible. Because the test is performed *in vivo*, it incorporates the ADME of the chemical into the response, furthering confidence that this assay is representative of organism-level effects. Antiandrogenic chemicals either bind to the AR and inhibit gene transcription or inhibit 5 α -reductase. These two MoAs differ in their effects on androgen-dependent tissues. For some chemicals, the different tissue effects identified in the Hershberger assay indicate the MoA (U.S. EPA, 2011b). In the castrated-peripubertal male rat, these tissues all respond to antiandrogens with a decrease in absolute weight (OECD, 2009a). Chemicals that increase steroid metabolism may also cause a decrease in the absolute weight of these organs. The EPA requires a statistically significant decrease in two or more endpoints to suggest antiandrogenic activity (U.S. EPA, 2011b). Confidence that the chemical has antiandrogenic potential increases as the number of significantly affected organs increases, and decreases as fewer endpoints respond. The Panel therefore assigned a concordant response in all Hershberger endpoints Rank 1 for the androgen antagonist hypothesis.

Rank 2 Endpoints

Binding affinity, as measured by competitive binding to the AR in the ARBA, was assigned Rank 2 status for the androgen antagonist hypothesis. AR binding *in vitro* suggests the potential for activity via this receptor. However, binding cannot predict the activation of transcription or distinguish whether the chemical has agonist, antagonist, or other effects via the receptor. Additional endocrine assays are required to determine whether the compound is an androgen antagonist (U.S. EPA, 2011c). The general limitations of *in vitro* assays, explained above, apply to the ARBA as well.

As stated above, significant growth responses in multiple androgen-dependent tissues (Cowper's gland, seminal vesicles, LABC, glans penis, and ventral prostate) in the antiandrogenic mode of the Hershberger assay suggests antiandrogenic potential (OECD, 2009a). The interpretation is most clear when several tissues respond concordantly. The EPA requires a statistically significant response from two or more of the tissues to interpret a positive response (U.S. EPA, 2011b), and therefore, the Panel assigned concordance of only two of the five endpoints Rank 2 for the androgen antagonist hypothesis.

In the pubertal male assay, Rank 2 status was given to age and weight, if increased, at PPS. Significantly increased age or weight (because they are older) at PPS in pubertal male rats can be a sensitive indicator of androgen antagonism because PPS is androgen-dependent. However, it is not exclusive to androgen antagonism. Chemicals that disrupt hypothalamic-pituitary function or steroid hormone synthesis or metabolism may also delay PPS (Stoker et al., 2000). Therefore, it is important to corroborate designation of an androgen antagonist through responses in other assays.

Reproductive organ weight, also assigned Rank 2 for androgen antagonism in the pubertal male assay, include weights of seminal vesicle and coagulating gland (wet and blotted), ventral prostate, dorsolateral prostate, LABC muscle complex, epididymis, and testes, as these are all dependent on androgens for growth (Shin et al., 2002). Although reduced reproductive organ weights are sensitive to androgen antagonism, these are not specific responses and not all organs may respond as expected. Androgen antagonists that chronically elevate serum LH concentrations (such as flutamide) will result in an increase in interstitial fluid volume of the Leydig cells and increase testes weight (Stoker et al., 2000; U.S. EPA, 2007b). Differences in the reproductive organ weight profiles for antiandrogens indicate that some androgen antagonists will not decrease weight in all of the reproductive organs utilized as endpoints in this assay. This confounds interpretation of the test and necessitates the use of other assays to determine whether the chemical is an androgen antagonist.

Histopathology of the testes and epididymides in the pubertal male assay were also assigned Rank 2 status for the androgen antagonist hypothesis. After treatment with the AR antagonist, flutamide, male pubertal rats displayed hyperplasia/hypertrophy of the interstitial cells in the testes and atrophy of the epididymides (Rocca and Pepperl, 2000). These effects are considered consistent with the androgen antagonist activity of chemicals that elevate serum LH concentrations. No decrease was detected in spermatogenesis. Testicular and epididymal histopathology is classified as a Rank 2 endpoint because they may be confounded by disease or overt systemic toxicity. The histopathology results are also dependent on secondary mechanisms of the androgen antagonists. For example, hyperplasia of interstitial cells in the testes may not be seen if the androgen antagonist does not increase serum LH concentrations.

A strength of the FSTRA is that it is an *in vivo* test, which incorporates the ADME of the chemical in the results and provides some direct measurements of reproductive organ physiology and function. Several endpoints in the FSTRA were assigned Rank 2 for the androgen antagonist hypothesis, including increased VTG in females, decreased secondary sex characteristics in males, and gonad histopathology in females. Plasma VTG concentrations may be, but are not always significantly increased in female fathead minnows exposed to antiandrogens (U.S. EPA, 2007a; Martinović et al., 2008). This increase in female VTG is likely based on a variety of mechanisms including direct interaction and antagonism with the AR. However, it may also be an indirect effect of reduced oocyte maturation, increased estradiol, or other mechanisms (Martinović et al., 2008). Therefore, while this effect is responsive to androgen antagonist exposure, overt toxicity or other nonendocrine effects confound interpretation of the endpoint. After a 30-day exposure to the androgen antagonist vinclozolin, male guppies exhibited a partial loss of orange-yellow coloration (a secondary sex characteristic of the species). Similarly, a concentration-dependent demasculinization of males (e.g., reduced expression of tubercles) upon treatment with vinclozolin has been observed with fathead minnows (Martinović et al., 2008). Although reduced secondary sexual characteristic

is suggestive of androgen antagonism, it is not diagnostic and is categorized as a Rank 2 endpoint. Alterations in gonad histopathology can be a sensitive indicator of endocrine dysfunction (U.S. EPA, 2007a). Exposure to vinclozolin altered gonad histopathology, including oocyte maturation and ovarian atresia in the early stages of vitellogenesis (U.S. EPA, 2007a; Martinović et al., 2008). In studies with the antiandrogen, flutamide, proliferation of Leydig cells, an increased number of spermatogonia, and an overall increase in testicular degeneration were observed in the testes (Jensen et al., 2004; U.S. EPA, 2007a). This endpoint was classified as Rank 2 because changes in gonad histopathology are not exclusively endocrine-mediated and may be the result of other methods of action or overt toxicity.

Rank 3 Endpoints

In the Hershberger assay, unique responses in weights of Cowper's gland, seminal vesicles, LABC, glans penis, and ventral prostate were assigned Rank 2 status. As stated above, a significant decrease in growth for two or more androgen-dependent tissues suggests a potential antiandrogenic MoA (U.S. EPA, 2011b). However, the situation is less clear if only one of the tissues shows a significant response. Because the EPA requires a statistically significant response from two or more of the tissues, one negative response is not considered sufficient evidence of antiandrogenic activity (U.S. EPA, 2011b). A single tissue with a significant decrease in growth is classified as a Rank 3 endpoint because it produces only qualitative evidence of antiandrogenic potential. A unique response should be evaluated with additional endpoints to determine the relevance of the response.

In the steroidogenesis assay, testosterone levels are assigned Rank 3 status for the androgen antagonist hypothesis. As stated above, the steroidogenesis assay is an *in vitro* assay utilized to detect endocrine effects mediated through the steroid hormone receptor. It is classified as a Rank 3 endpoint because it is not specific to androgen effects. Contrary to *in vivo* results in rats (see below), the androgen antagonist vinclozolin significantly decreased *in vitro* testosterone levels in the steroidogenesis assay (Hecker et al., 2006). This suggests that the steroidogenesis pathway does not mediate the antiandrogenic effects of vinclozolin. Because this assay is not specific for androgen antagonists, it should be utilized only as corroboration of results from Ranks 1 and 2 endpoints and should not be the determining factor for confirming or refuting the hypothesis.

Testosterone levels in the pubertal male assay are also assigned Rank 3 status for the androgen antagonist hypothesis. Androgen antagonists increase serum testosterone levels. The increase in serum testosterone is hypothesized to be a result of increased serum LH concentrations. Androgen antagonists that bind to the AR competitively inhibit the binding of endogenous androgens. This reduces the androgenic signal to the hypothalamus and increases LH secretion, which stimulates testosterone production in the Leydig cells (Shin et al., 2002). An increase in testosterone levels has several possible causes and is not interpretable in isolation.

However, this endpoint is helpful when assessed with results from higher ranked endpoints.

Fecundity, estradiol and testosterone levels, GSI, behavior, and fertilization success in the FSTRA were given Rank 3 status for the androgen antagonism hypothesis. Reduced fecundity is observed after exposure to androgen antagonists. While this endpoint is consistent and reproducible, it is not specific. Because fecundity involves both sexes, it may decrease whether the chemical is an estrogen agonist, estrogen antagonist, androgen agonist, or androgen antagonist (U.S. EPA, 2007a). Additionally, fecundity can be influenced by a variety of nonendocrine factors. Therefore, it can support a determination, but does not indicate androgen antagonism by itself. Androgen antagonists usually increase circulating levels of estradiol and testosterone, however the response of sex steroid levels are not entirely predictable in the antiandrogen pathway (OECD, 2004a; U.S. EPA, 2007a). As with other endpoints, changes in sex steroid levels may be caused by non-endocrine-mediated MoAs. They may be useful, however, in providing additional evidence that the observed effects are endocrine-mediated and are often used to support an endocrine-mediated MoA for decreases in fecundity (U.S. EPA, 2007a). GSI is an indirect measure of reproductive readiness. The preliminary assumption for this endpoint is that the relationship between gonad and body weight is constant. However, this assumption does not hold true during spawning, indicating this endpoint cannot be interpreted without corroboration from Ranks 1 and 2 endpoints (OECD, 2004a). After a 30-day exposure to the androgen antagonist vinclozolin, male guppies exhibited a decreased GSI (OECD, 2004a). However, alterations in GSI were variable following different antiandrogenic exposures with the fathead minnow (U.S. EPA, 2007a). In zebrafish, antiandrogenic exposure reduces postdawn spawning behavior (OECD, 2009b). As stated previously, behavior endpoints are subjective and may be caused by a variety of nonendocrine mechanisms. Therefore, this endpoint should not be utilized as a sole means of determining potential interaction with an endocrine pathway. Finally, exposure to androgen antagonists can significantly decrease the number of embryos that hatch, even in the absence of overt toxicity (Jensen et al., 2004). An endocrine-related reduction in fertilization could be due to elevated levels of atresia in the female gonad and decreased spermatogenesis in males. Because fertilization success is not specific to androgen antagonist exposures, this endpoint was classified as Rank 3.

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN AGONIST WITH COMPONENTS OF THYROID PATHWAYS

Thyroid hormone is essential for normal development, growth, neural differentiation, and metabolic regulation in mammals and for metamorphosis in amphibians (Yen, 2001; Brent, 2012). The requirement for thyroid hormone is most apparent when the hormone is deficient during development, such as in maternal iodine deficiency or untreated congenital hypothyroidism. While developmental deficiencies produce profound neurologic deficits and growth retardation (Yen, 2001), more subtle and reversible

Table 5
Thyroid Agonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
AMA Asynchronous development Thyroid histopathology	Pubertal male assay Thyroid weight TSH levels T4 levels Pubertal female assay TSH levels T4 levels AMA Advanced developmental stage Hind limb length: increased	Pubertal male assay Growth Age and weight at preputial separation Pituitary weight Pubertal female assay Growth Age and weight at vaginal opening Blood chemistry AMA Snout-vent length: reduced Wet weight: reduced

defects may develop when ligand deficiency occurs in the adult (Brent, 2012). The effects of thyroid hormone are complex, involving the interactions of T_4 , T_3 , the two main isoforms of thyroid receptors, $TR\alpha$ and $TR\beta$, each of which has two subtypes, and thyroid-stimulating hormone (TSH) (Brent, 2012). Hepatic metabolism and clearance of thyroid hormones is also critical for control of this pathway. Interaction of thyroid hormone with specific receptors in various tissues regulates transcription of a wide variety of genes, initiates nongenomic cellular effects, and modulates the cellular responses to a variety of other hormones (Yen, 2001; Brent, 2012). Because of its role in growth and metabolism and its involvement in the activities of other hormones, particularly the sex steroid hormones, the identification of specific hallmark mammalian responses, as has been done for androgens and estrogens, is very difficult for thyroid hormone. The metamorphic response to thyroid hormone is somewhat more specific in amphibians (Galton, 1992; Tata, 2006). Table 5 lists proposed Ranks 1, 2, and 3 endpoints for the thyroid agonist hypothesis.

Rank 1 Endpoints

Rank 1 endpoints for the thyroid agonist hypothesis are asynchronous development and thyroid histopathology assessed in the AMA. Thyroid hormone is an obligatory signal for the initiation and completion of amphibian metamorphosis and its effects on amphibian tissue are direct, local, and independent of the location of the target cells (Tata, 2006). Normal metamorphic development requires precise synchronization of the increase in thyroid hormone production and release (Galton, 1992; Tata, 2006), hence, asynchronous metamorphic development is a reliable indication of potential thyroid agonist activity and is given Rank 1 priority. Furthermore, it has been the Panel's experience that there is a low to no occurrence of asynchronous development in control tadpoles in the AMA, giving weight to the understanding that this observation is likely specific to alterations in the HPT axis. Xenobiotic chemicals capable of thyroid hormone activity appear to primarily affect processes related to thyroid

hormone synthesis and iodine uptake and incorporation into thyroid hormone, and less prominently to interaction via thyroid hormone receptors (Degitz et al., 2005). Therefore, an *in vivo* assay is required to capture predominant modes of potential thyroid agonist (or antagonist) activity.

Unlike most other endocrine glands, the thyroid gland stores the hormone it produces extracellularly, in follicles. Thyroid hormone production during development and metamorphosis results in changes in thyroid histology, forming the basis for prioritizing this endpoint as Rank 1. The more frequently examined histological features used to measure thyroid activity in metamorphosing taxa are gland volume, follicle volume, number of colloid-filled follicles, and epithelial cell height (Jennings and Hanken, 1998). Thyroid gland hypertrophy, thyroid gland atrophy, follicular cell hypertrophy, and follicular cell hyperplasia are specified as core thyroid histopathology parameters examined in the AMA (Grim et al., 2009). It is important to note that the normal histomorphology of the thyroid gland changes as development of the tadpole progresses. Thus, these normal changes must be considered when thyroid histology of tadpoles in different developmental stages is compared (Grim et al., 2009). Exposure to the endogenous thyroid agonist, T_4 , resulted in thyroid gland atrophy among African-clawed frog tadpoles in the AMA (Coady et al., 2010), indicating this response is specific to agonists of the thyroid pathway.

Rank 2 Endpoints

In the pubertal male assay, thyroid weights and circulating T_4 and TSH levels are proposed as Rank 2 endpoints. Although interpretable without corroboration from other endpoints in the battery, the rationale for including these endpoints in Rank 2, rather than in Rank 1, is due to normal variability in thyroid weights and T_4 and TSH levels and the potential for these endpoints to be affected by systemic toxicity and stress. Thyroid hormone measurements have large coefficients of variation (27 and 58% for T_4 and TSH, respectively; U.S. EPA, 2009), and are subject to rapid, stress-induced changes as may occur during necropsy. Thyroid weights have been found to be highly variable and often outside the maximum and minimum acceptable control limits set forth in the test guideline (EPF, 2013). Hepatic toxicity and associated liver enzyme induction may also produce indirect effects on thyroid hormones due to alterations in clearance and compensatory homeostatic changes.

In the pubertal female assay, T_4 and TSH levels are proposed as Rank 2 endpoints. As with thyroid hormone measurements in the pubertal male assay, TSH and T_4 levels can be variable and influenced by nonendocrine factors; thus, the interpretation is not as clear as required for Rank 1 status.

Advanced development, as measured by advanced developmental stage or increased normalized hind limb length relative to controls in the AMA is also assigned Rank 2. Although an expected effect of thyroid agonists, experience among the Panel indicates that interpretation of advanced development requires information on thyroid histopathology for both thyroid agonists and antagonists. Furthermore, advanced development is defined by

either increased normalized hind limb length or advanced developmental stage relative to controls, but it is not necessary for both to be present for determination of a thyroid agonist.

Rank 3 Endpoints

Rank 3 endpoints in the pubertal male assay for the thyroid agonist hypothesis include growth, age, and weight at PPS and pituitary weight. The interpretability of the growth and age at PPS has been addressed generally and for other hypotheses previously. Pituitary weights have been found to be highly variable (EPF, 2013) and not interpretable without other markers of thyroidal effects in the pubertal female assay, also given Rank 3, which include growth, age and weight at VO and blood chemistry. The interpretability of these endpoints has also been discussed generally and for other hypotheses previously.

Alterations in growth in the AMA are considered to be Rank 3 endpoints for the thyroid agonist hypothesis. Reductions in tadpole growth (as measured by decreased wet weight and snout-vent length) have been coupled with advanced development rate as the growth phase of tadpole development is truncated. However, reduced tadpole growth, by itself, should not be relied upon to determine thyroid-specific agonism, since reduction in tadpole growth is a response that is subject to various modes of toxic action and ecological influences that are not specific to interactions with the HPT axis (U.S. EPA, 2007b).

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN ANTAGONIST WITH COMPONENTS OF THYROID PATHWAYS

Antagonists of thyroid hormone can profoundly retard normal metabolism and development, primarily via changes in thyroid hormone production or clearance (Degitz et al., 2005). Altered production and clearance of thyroid hormones is theoretically most evident in direct pathological changes or compensatory responses of the thyroid gland. As such, mammalian assays are relatively more useful in the detection of thyroid antagonists than agonists. Table 6 lists proposed Ranks 1, 2, and 3 endpoints for the thyroid antagonist hypothesis.

Rank 1 Endpoints

As in the case of the thyroid agonist hypothesis, asynchronous development (especially for peripheral alterations in deiodinase activity) and thyroid histopathology in the AMA are Rank 1 endpoints for the thyroid antagonist hypothesis. Previous research has indicated that the histopathological examination of the thyroid gland is a sensitive and specific indicator of thyroid activity, particularly for compounds that inhibit thyroid hormone synthesis in the AMA (Degitz et al., 2005; Coady et al., 2010; Pickford, 2010). Thyroid weight increase and histopathology in the pubertal male assay (Marty et al., 2001a), and thyroid weight and colloid area and follicular cell height in the female pubertal assay are reliably affected by thyroid antagonists. In mammals, alterations

in thyroid histopathology are quantified as changes in follicular cell height and volume, with follicular cell hypertrophy and hyperplasia the most common responses to decreased circulating T_4 and increased TSH. Although this is considered a Rank 1 endpoint, it should be noted that the rat overpredicts thyroid toxicity in man, due to its chronic “stimulated” state resulting from the relative weak binding of the circulating thyroid hormones (Jahnke et al., 2004).

Rank 2 Endpoints

For the thyroid antagonist hypothesis, liver weight, T_4 and TSH levels in the pubertal male assay (Marty et al., 2001a; Stoker and Zorrilla, 2010), and age and weight at VO, T_4 and TSH levels in the pubertal female assay (Marty et al., 1999; Stoker and Zorrilla, 2010) are given Rank 2 priority. The limitations of these endpoints relative to Rank 1 endpoints have been discussed generally above and for other hypotheses (Marty et al., 2001a, 2001b). Testicular weight, histopathology, and sperm counts (if conducted) in the pubertal male assay also may respond to thyroid hormone depletion postnatally as Sertoli cell production may be altered and testis weight increased during development and sperm counts increase (Cooke et al., 1993, 1996). This finding is more likely to be captured during a Tier II reproductive toxicity study, although the pubertal male animals are just past weaning at study initiation and might still be susceptible.

Rank 3 Endpoints

The Panel assigned growth, age and weight at PPS, and pituitary weights in the pubertal male assay and estrous cyclicity and ovary histopathology with reduced ovary weight or atrophy in the pubertal female assay as Rank 3 endpoints. The limitations of these various endpoints have been discussed generally above (Marty et al., 2001a, 2001b). In the pubertal female assay, high doses of thyroid antagonists are expected to produce irregular cyclic-

ity (diestrus) and may induce ovarian atrophy (Hagino 1976; Armada-Dias et al., 2001; York et al., 2001; Hatsuta et al., 2004; Hapon et al., 2010; Li et al., 2011).

Thyroid antagonism increases tadpole growth because the tadpoles do not progress along the expected timeline for metamorphic reorganization and tissue loss; however, alteration in growth (both increased and decreased) is an apical endpoint that can respond to many different influences apart from the thyroid pathway (OECD, 2004b; 2007c; 2009c). Thus, growth changes in the tadpole can only be used as Rank 3 to support more definitive endpoint responses for the thyroid antagonist hypothesis. Likewise, when morphological signs of delayed development (such as reduced hind limb length or delayed developmental stage) suggest a test chemical may have potential antagonism in the thyroid pathway, this conclusion needs to be corroborated by more definitive endpoints such as thyroid gland histopathology, because delays in tadpole development can be the result of delayed growth rather than the result of specific perturbations in the HPT axis (Wilbur and Collins, 1973; U.S. EPA, 2007b).

HYPOTHESIS: THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN INDUCER OF STEROIDOGENIC ENZYMES

Steroidogenesis entails a process whereby cholesterol is converted to biologically active steroid hormones. This can be understood as a single process that is repeated, with cell-specific variations, in each endocrine gland (Miller and Auchus, 2011). Steroidogenesis is rate limited at conversion of cholesterol to pregnenolone by CYP11A1, which is the initial step in the process. Although mediated by a single enzyme, this initial step is enzymatically complex and subject to multiple regulatory mechanisms that enable precise quantitative control over the rate of conversion (Miller and Auchus, 2011). Qualitative control over the type of steroid produced is governed by the cellular content of the two classes of enzymes involved in steroidogenesis, CYP-dependent mixed function oxidases, and hydroxysteroid dehydrogenases (Miller and Auchus, 2011). Both classes of enzymes are relevant molecular targets for endocrine activity due to their role in maintaining the androgen/estrogen hormonal balance. Interactions between chemicals and critical steroidogenic enzymes may alter this balance, and hence, alter endocrine functioning (Quignot et al., 2012). Aromatase (CYP19) is a critical enzyme in the steroidogenic pathway and is principally responsible for the conversion of 19C-androgens to 18C-estrogens (Marty et al., 2001b; Quignot et al., 2012) and specific inhibitors of the enzyme are well characterized (Hecker et al., 2011).

The steroidogenesis induction hypothesis, as well as the steroidogenesis inhibition hypothesis (Section 8), are formulated differently in the present report compared with the Framework publication (Borgert et al., 2011a). In the Framework, interaction with aromatase was considered in separate hypotheses from interactions with steroidogenic enzymes. With the experience gained from presenting this scheme at numerous scientific meetings after the Framework was published and from conducting the Tier 1 ESB on dozens of chemicals, the Panel determined that

Table 6
Thyroid Antagonist Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
Pubertal male assay	Pubertal male assay	Pubertal male assay
Thyroid weight	Liver weight	Growth
Thyroid histopathology	T_4 levels	Age and weight at prepubertal separation
	TSH levels	Pituitary weight
Pubertal female assay	Pubertal female assay	AMA
Thyroid weight	Age and weight at vaginal opening	Delayed development
Thyroid (colloid area and follicular cell height)	T_4 levels	Snout-vent length
	TSH levels	Hind limb length
		Wet weight
AMA		Pubertal female assay
Asynchronous development		Estrous cyclicity (diestrus)
Thyroid histopathology		Ovary histopathology: reduced ovary weight or atrophy

Table 7
Steroidogenesis Induction Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
	Steroidogenesis assay	Male pubertal assay
	Estradiol levels	Testosterone levels
	Testosterone levels	Testosterone levels

it is more relevant to formulate these hypotheses with respect to the potential of a chemical to interrupt steroidogenesis, of which aromatase should be considered a component. This revised formulation is also more consistent with the formulation of other hypotheses according to agonist and antagonist modalities. According to the EPA's EDSP guidance document (U.S. EPA, 2011a), five Tier 1 assays are capable of detecting an alteration in steroidogenesis due to chemical exposure: steroidogenesis, aromatase, pubertal male, pubertal female, and FSTRA. The aromatase and pubertal female evaluate estrogenic, the pubertal male evaluates androgenic, and the steroidogenesis assay and FSTRA evaluate both estrogenic and androgenic MoAs, both of which may be affected after significant alterations in steroidogenesis. Table 7 lists proposed Ranks 1, 2, and 3 endpoints for the steroidogenesis inducer hypothesis.

Rank 1 Endpoints

No *in vivo* endpoints were considered to be sufficiently specific and interpretable to warrant Rank 1 status for the potential induction of steroidogenic enzymes. This limitation arises, in part, because of the lack of clear inducers of steroidogenesis that could be used to verify responses in specific *in vivo* endpoints.

Rank 2 Endpoints

In the steroidogenesis assay, testosterone levels are the central product and intermediate in the production of estrogens and hence, measurement of this endpoint is assigned Rank 2 status, consistent with the Panel's principle that only *in vivo* assays should be assigned Rank 1. However, this assay is the most specific of the Tier 1 battery for interaction with the steroidogenic pathway despite that it is subject to the limitations of all *in vitro* assays. Accordingly, altered testosterone levels are expected after exposures to compounds that alter the steroidogenesis enzyme system. As the ultimate product of the steroidogenesis pathway, estradiol levels are also assigned Rank 2. Additional Rank 2 endpoints are not identifiable due to the lack of chemicals with known steroidogenic induction activity.

Rank 3 Endpoints

The lack of chemicals with known steroidogenic induction activity limits the endpoints that can be assigned relevance for this hypothesis, but it might be expected that steroid hormone levels would be increased *in vivo* by strong inducers of steroidogenesis. Therefore, testosterone levels in the male pubertal assay were assigned Rank 3 for induction of steroidogenesis.

Table 8
Steroidogenesis Inhibition Hypothesis

Rank 1 endpoints	Rank 2 endpoints	Rank 3 endpoints
Pubertal female assay	Steroidogenesis assay	Aromatase assay
Uterus weight	Estradiol levels	Aromatase inhibition
	Testosterone levels	
		Pubertal female assay
	Pubertal female assay	Age and weight at vaginal opening: increased
	Ovary weight	Age at first estrus: increased
	FSTRA	
	Vitellogenin: reduced in females	
	Gonad histopathology: males	FSTRA
		Fecundity
		Estradiol levels
		Testosterone levels
		Gonad-somatic index
		Behavior
		Fertilization success

THE CHEMICAL EXHIBITS THE POTENTIAL TO INTERACT AS AN INHIBITOR OF STEROIDOGENESIS ENZYMES

The brief description of steroidogenesis provided under the steroidogenesis induction hypothesis (Section 7) applies also to the steroidogenesis inhibitor hypothesis. Table 8 lists proposed Ranks 1, 2, and 3 endpoints for the steroidogenesis inhibitor hypothesis.

Rank 1 Endpoints

In the pubertal female assay, uterine weights are significantly reduced by potent aromatase inhibitors, although there is some concern about how effectively this endpoint would respond to substances with weak aromatase inhibition capacity (Marty et al., 1999). Because this *in vivo* endpoint measures an apical response encompassing ADME and allows an estimate of strength of response, the Panel assigned significant reduction in uterine weight in the pubertal female assay Rank 1 priority for the steroidogenic inhibitor hypotheses.

Rank 2 Endpoints

In the steroidogenesis assay, testosterone levels are the central product and intermediate in the production of estrogens and hence, measurement of this endpoint is assigned Rank 2 status, consistent with the Panel's principle that only *in vivo* assays should be assigned Rank 1. However, this assay is the most specific of the Tier 1 battery for interaction with the steroidogenic pathway despite that it is subject to the limitations of all *in vitro* assays. Accordingly, altered testosterone levels are expected of compounds that interfere with the steroidogenesis enzyme system. As the ultimate product of the steroidogenesis pathway, estradiol levels are also assigned Rank 2 endpoint.

In the pubertal female assay, ovarian weight is sensitive to perturbation of the steroidogenic enzyme system and is expected to respond to chemicals with such potential (Marty et al., 2001b; Stoker and Zorrilla, 2010). As such,

ovarian is assigned Rank 2 for the steroidogenesis inhibition hypothesis.

In the FSTRA, a decrease in female VTG and gonadal histopathology, particularly in male fish, is considered a Rank 2 endpoint for the steroidogenesis inhibition hypothesis. In several investigations, VTG was decreased in female fathead minnows in response to inhibitors of steroidogenesis, while there was no detectable change in male VTG levels in response to inhibitors of steroidogenesis (U.S. EPA, 2007a). Reductions in female VTG are not considered to be Rank 1 endpoints, since other stresses and other toxicity pathways can result in reduced VTG levels (Anderson et al., 1996a, 1996b; Lethimonier et al., 2000; Celander, 2011; Bugel et al., 2013). Nonetheless, reduced female VTG in female fish would be expected to be a fairly sensitive response for detecting inhibition of steroidogenesis. Among male fathead minnows, testicular degeneration, enlarged seminiferous tubule lumen, proliferation of Leydig cells, and increased spermatozoa were observed when fish were exposed to steroidogenesis inhibitors (Ankley et al., 2007; U.S. EPA, 2007a). These responses were observed in several independent laboratories and thus support the use of male gonadal histopathology as a Rank 2 endpoint in the steroidogenesis inhibition pathway.

Rank 3 Endpoints

Aromatase inhibition in the aromatase assay is considered a Rank 3 endpoint for the potential to interact with the steroidogenic enzyme system. It is not given a higher status primarily because it comprises only a single component of the system. However, it should be considered a strong corroborative endpoint and should help to clarify whether the point of interaction is aromatase (CYP19) for compounds that produce responses in Ranks 1 and 2 endpoints.

In the pubertal female assay, age and weight at VO and age at first estrus are increased by potent aromatase inhibitors, although the response may be less reliable than for uterine weight reductions (Marty et al., 1999).

In the FSTRA, fecundity, sex steroid levels, GSI, behavior, and fertilization success were classified as Rank 3 because they are relevant, but not specific to inhibitors of steroidogenesis. Reduced fecundity was observed in response to inhibitors of steroidogenesis in the FSTRA (U.S. EPA, 2007a). Also, male GSI was increased and there was either no change or a decrease in female GSI in response to steroidogenesis inhibitors in the FSTRA (Ankley et al., 2007; U.S. EPA, 2007a). Therefore, increased male GSI may be informative to support the hypothesis of steroidogenesis inhibitors in conjunction with more definitive Ranks 1 and 2 endpoints. Likewise, sex steroid concentrations in the FSTRA can be informative as Rank 3 endpoints. When exposed to inhibitors of steroidogenesis, the general profile of sex steroid responses was a decrease in female estradiol and an increase in female testosterone, while male fish generally exhibited an increase in testosterone and 11-ketotestosterone (U.S. EPA, 2007a). These endpoints are considered Rank 3 due to their lower overall sensitivity, relatively greater variability in response, and their ability to be influenced by other modes of action apart from just steroidogenesis inhibition.

DISCUSSION

This work provides justifications for the assignment of relevance weight (W_{REL}) rankings to experimental endpoints from the U.S. EPA's ESB according to their relevance for deciding eight specific hypotheses and an interpretive process by which the rankings are used to reach a WoE conclusion. Because data do not exist for weighting the endpoints based on positive and negative predictive values, the rankings proposed here were based on the collective scientific judgment of an expert panel of scientists from the EPF (the Panel) as informed by their experience with the assays and the relevant scientific literature cited herein. Although these W_{REL} rankings are not quantitatively derived, their a priori explication nonetheless enhances transparency and renders WoE determinations amenable to methodological scrutiny.

There are several advantages to weightings and interpretive criteria derived and explained a priori, as offered here, rather than as used to explain WoE determinations post hoc. A priori rankings and interpretive criteria provide a transparent set of assumptions that can be used to conduct a WoE evaluation for any particular chemical, and thus, different analysts should derive similar WoE conclusion based on the same data. This helps to assure that WoE determinations are clear and consistent from chemical to chemical, an aspect that is particularly important for substances subjected to the EDSP ESB due to the high visibility of the program.

Clearly defined weightings and interpretive criteria also allow WoE determinations to be subjected to methodological scrutiny rather than to subjective debates about judgments. If different conclusions are reached by different analysts, the methodology provides a means of readily identifying the source of the difference, as the underlying assumptions of the method are defined a priori. WoE conclusions based on the Framework, including the W_{REL} rankings and interpretive criteria proposed here, can be examined for sensitivity to the underlying assumptions of the method. For example, the rankings for particular endpoints could be reassigned to determine how their assessment affects the overall support for any hypothesized endocrine activity. This could be useful for judging the strength of the proposed rankings as well as for providing a means for individual analysts to test rankings that appear questionable or incorrect. This seems a particularly important strength, since different scientific judgments could be made for many of the proposed rankings. The ability to change the rankings and to include additional endpoints in specific ranked categories also provides a ready means for the methodology to incorporate new scientific information as it emerges.

The disadvantage of the relevance weightings and interpretive criteria proposed here is their lack of quantitative foundation. Although this does not diminish the advantages discussed above, it must be kept in mind that the EDSP was to be reviewed and reevaluated following the completion of an initial screening of 50 to 100 chemicals (U.S. EPA, 1999). A complete review of the data is needed and should include an evaluation of individual assay performance as well as an evaluation of the positive and negative predictive value of the battery as a whole for reproductive and developmental toxicity. Such a review

and evaluation would provide a more quantitative basis for adjusting the methodology proposed here. Although an evaluation of positive and negative predictive values would not satisfy causality requirements for demonstrating that a particular adverse effect is mediated by an endocrine mode of action, such information would aid in determining the effectiveness and efficiency of the EDSP for protecting public health and environment.

This article uses the ESB assays to illustrate a process and rationale for transparently ranking endpoints for WoE evaluations. Although the process is broadly applicable, we do not portend to include all studies that might be used to assess potential endocrine activity. While such a treatise would be useful, it is beyond the scope of a single publication. Nonetheless, it is hoped that analysts addressing other endpoints and other hormonal modalities could apply the principles and rationale illustrated here to develop objective and transparent relevance rankings.

Identifying studies that might be used in WoE assessments requires careful consideration of the goals of the particular assessment and how unambiguously each endpoint would inform those goals. Several different WoE evaluations are implied by the U.S. EPA's EDSP (Borgert et al., 2011a), and different types of studies may be required for each. If the goal of the assessment is to identify the potential to act via specific endocrine modes of action, as is the intent of the ESB, then it is important to select endpoints with a high degree of mechanistic specificity. Many long-term studies measure apical endpoints that could be affected by endocrine activity, but which could also be affected by other types of toxicity. Even though such endpoints may be quite relevant for assessing adverse effects, they may have little utility for identifying a mode of action. Conversely, endpoints with sufficient mechanistic specificity to identify potential modes of action may not be definitive for adverse effects and hence, for WoE evaluations intended to inform risk assessments.

CONFLICT OF INTEREST STATEMENT AND FUNDING DISCLOSURE

Some substances under EPA EDSP Test Orders for endocrine screening are produced by the authors' employers. This article has been reviewed in accordance with the peer- and administrative-review policies of the authors' organizations. The views expressed here are those of the authors and do not necessarily reflect the opinions and/or policies of their employers. There are no contractual relations or proprietary considerations that restrict dissemination of the research findings of the authors. C. J. Borgert, E. M. Mihaich, and L. D. Stuchal are independent scientists/consultants who received financial support for portions of this project from the Endocrine Policy Forum. Time spent by other co-authors was supported by their respective employers or a personal contribution.

REFERENCES

Aluru N, Vijayan MM. 2009. Stress transcriptomics in fish: a role for genomic cortisol signaling. *Gen Comp Endocrinol* 164:142–150.

American Heritage® Medical Dictionary Copyright©. 2007 by Houghton Mifflin Company. Published by Houghton Mifflin Company, Boston, MA.

Anderson MJ, Miller MR, Hinton DE. 1996a. *In vitro* modulation of 17- α -estradiol-induced vitellogenin synthesis: effects of cytochrome

P4501A1-inducing compounds on rainbow trout (*Oncorhynchus mykiss*) liver cells. *Aquat Toxicol* 34:327–350.

Anderson MJ, Olsen H, Matsumura F, Hinton DE. 1996b. *In vivo* modulation of 17- α -estradiol-induced vitellogenin synthesis and estrogen receptor in rainbow trout (*Oncorhynchus mykiss*) liver cells by β -naphthoflavone. *Toxicol Appl Pharmacol* 137:210–218.

Ankley GT, Gray LE. 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(5):1084–1087.

Ankley GT, Villeneuve DL. 2006. The fathead minnow in aquatic toxicology: past, present and future. *Aquat Toxicol* 78:91–102.

Ankley G, Mihaich E, Stahl R, Tilitt D, Colborn T, McMaster S, Miller R, Bantle J, Campbell P, Denslow N. 1998. Overview of a workshop on screening methods for detecting potential (anti-) estrogenic/androgenic chemicals in wildlife. *Environ Toxicol Chem* 17: 68–87.

Ankley GT, Jensen KM, Kahl MD, Korte JJ, Makynen EA. 2001. Description and evaluation of a short-term reproduction test with the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1276–1290.

Ankley GT, Jensen KM, Makynen EA, Kahl MD, Korte JJ, Hornung MW, Henry TR, Denny JS, Leino RL, Wilson VS, Cardon MC, Hartig PC, Gray LE. 2003. Effects of the androgenic growth promoter 17- β -trenbolone on fecundity and reproductive endocrinology of the fathead minnow. *Environ Toxicol Chem* 22:1350–1360.

Ankley GT, Jensen KM, Kahl MD, Makynen EA, Blake LS, Greene KJ, Johnson RD, Villeneuve DL. 2007. Ketoconazole in the fathead minnow (*Pimephales promelas*): reproductive toxicity and biological compensation. *Environ Toxicol Chem* 26:1214–1223.

Ashby J, Lefevre PA. 2000. The peripubertal male rat assay as an alternative to the Hershberger castrated male rat assay for the detection of anti-androgens, oestrogens and metabolic modulators. *J Appl Toxicol* 20:35–47.

Armada-Dias L, Carvalho J, Breitenbach M, Franci C, Ooura E. 2001. Is the fertility in hypothyroidism mainly due to ovarian or pituitary functional changes? *Braz J Med Biol Res* 34:1209–1215.

Armstrong EJ, Federman DD. 2005. Pharmacology of reproduction, chapter 27. Principles of pharmacology: the pathophysiologic basis of drug therapy. In: Golan DE, editor. Philadelphia, PA: Lippincott Williams & Wilkins.

Ashby J, Owens W, Deghenghi R, Odum J. 2002. Concept evaluation: an assay for receptor-mediated and biochemical antiestrogens using pubertal rats. *Regul Toxicol Pharmacol* 35:393–397.

Baker ME. 2002. Beyond carrier proteins albumin, steroid hormones and the origin of vertebrates. *J Endocrinol* 175:121–127.

Battelle. 2005. Final detailed review paper on aromatase. EPA Contract 01–023.

Beri R, Kumar N, Savage T, Benalcazar L, Sundaram K. 1998. Estrogenic and progestational activity of 7 α -methyl-19-nortestosterone, a synthetic androgen. *J Steroid Biochem Mol Biol* 67:275–283.

Black LJ, Jones CD, Goode RL. 1981. Differential interaction of antiestrogens with cytosol estrogen receptors. *Mol Cell Endocrinol* 22:95–103.

Borgert CJ, Mihaich EM, Ortego LS, Bentley KS, Holmes CM, Levine SL, Becker RA. 2011a. Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disruptor Screening Program. *Regul Toxicol Pharmacol* 61:185–191.

Borgert CJ, Mihaich EM, Quill TF, Marty MS, Levine SL, Becker RA. 2011b. Evaluation of EPA's Tier 1 Endocrine Screening Battery and recommendations for improving the interpretation of screening results. *Regul Toxicol Pharmacol*. 59:397–411.

Borgert CJ, Sargent EV, Casella G, Dietrich DR, McCarty LS, Golden RJ. 2012. The human relevant potency threshold: reducing uncertainty by human calibration of cumulative risk assessments. *Regul Toxicol Pharmacol* 62:313–328.

Borgert CJ, Baker SP, Matthews JC. 2013. Potency matters: thresholds govern endocrine activity. *Regul Toxicol Pharmacol* 67:83–88.

Bovee TFH, Pikkemaat MG. 2009. Bioactivity-based screening of antibiotics and hormones. *J Chromatogr A* 1216:8035–8050.

Brent GA. 2012. Mechanisms of thyroid hormone action. *J Clin Invest* 122:3035–3043.

Bugel SM, White LA, Cooper KR. 2013. Inhibition of vitellogenin gene induction by 2,3,7,8-tetrachlorodibenzo-p-dioxin is mediated by aryl hydrocarbon receptor 2 (AHR2) in zebrafish (*Danio rerio*). *Aquat Toxicol* 126:1–8.

Bulger WH, Kupfer D. 1985. Estrogenic activity of pesticides and other xenobiotics on the uterus and male reproductive tract. *Endocrine toxicology*. Dixon, RL., Thomas, JA, Korach, KS, McLachlan JA., Eds. New York: Raven Press.

- Caldwell DJ, Mastrocco F, Anderson PD, Länge R, Sumpter JP. 2012. Predicted-no-effect concentrations for the steroid estrogens estrone, 17 β -estradiol, estriol, and 17 α -ethynylestradiol. *Environ Toxicol Chem* 31:1396–1406.
- Celander MC. 2011. Cocktail effects on biomarker responses in fish. *Aquat Toxicol* 105S:72–77.
- Chapin RE, Gulati DK, Barnes LH, Teague JL. 1993. The effects of feed restriction on reproductive function in Sprague-Dawley. *Fundam Appl Toxicol* 20, 23–29.
- Clark AS, Kelton MC, Whitney AC. 2003. Chronic administration of anabolic steroids disrupts pubertal onset and estrous cyclicity in rats. *Biol Reprod* 68:465–471.
- Clearwater SJ, Pankhurst NW. 1997. The response to capture and confinement stress of plasma cortisol, plasma sex steroids and vitellogenic oocytes in the marine teleost, red gurnard. *J Fish Biol* 50:429–441.
- Coady K, Marino T, Thomas J, Currie RJ, Hancock G, Crofoot J, McNalley L, McFadden L, Getter D, Klecka G. 2010. Evaluation of the amphibian metamorphosis assay: exposure to the goitrogen methimazole and the endogenous thyroid hormone L-thyroxine. *Environ Toxicol Chem* 29:869–880.
- Cooke PS, Kirby, JD, Porcelli, J. 1993. Increased testis growth and sperm production in adult rats following transient neonatal goitrogen treatment: optimization of the propylthiouracil dose and effects of methimazole. *J Reprod Fertil* 97(2):493–499.
- Cooke PS, Zhao YD, Hansen LG. 1996. Neonatal polychlorinated biphenyl treatment increases adult testis size and sperm production in the rat. *Toxicol Appl Pharmacol* 136(1):112–117.
- Degitz SJ, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, Tietge JE. 2005. Progress towards development of an amphibian-based thyroid screening assay using *Xenopus laevis*. Organismal and thyroidal responses to the model compounds 6-propylthiouracil, methimazole, and thyroxine. *Toxicol Sci* 87:353–364.
- Endocrine Policy Forum (EPF). 2013. Available at: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2013-0230-0020>. Accessed October 25, 2013.
- Foo JTW, Lam TJ. 1993. Retardation of ovarian growth and depression of serum steroid levels in the tilapia, *Oreochromis mossambicus*, by cortisol implantation. *Aquaculture* 115:133–143.
- Fritsch MK, Murdoch FE. 1998. Estrogens, progestins, and contraceptives. Chapter 38. Human pharmacology: molecular to clinical. In: Brody TM, editor. St. Louis, MO: Mosby-Year Book, Inc.
- Galton VA. 1992. The role of thyroid hormone in amphibian metamorphosis. *Trends Endocrinol Metab* 3:96–100.
- Goldman JM, Laws SC, Balchak SK, Cooper RL, Kavlock RJ. 2000. Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. *Crit Rev Toxicol* 30:135–196.
- Gori GB. 2010. Regulating unknown risk. *Regulation* 33:16–21.
- Gracia T, Hilscherova K, Jones PD, Newsted JL, Higley EB, Zhang X, Hecker M, Murphy MB, Yu RM, Lam PK, Rudolf RS, Giesy JP. 2007. Modulation of steroidogenic gene expression and hormone production of H295R cells by pharmaceuticals and other environmentally active compounds. *Toxicol Appl Pharmacol* 225:142–153.
- Gray LE, Wilson V, Noriega N, Furr J, Stoker TE, Laws SC, Goldman J, Cooper RL, Foster PMD. 2004. Use of the laboratory rat as a model in endocrine disruptor screening and testing. *ILAR J* 45:425–437.
- Grim KC, Wolfe M, Braunbeck T, Iguchi T, Ohta Y, Tooi O, Touart L, Wolfe DC, Tietge J. 2009. Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active substances. *Toxicol Pathol* 37:415–424.
- Guyton Arthur C., John E. Hall. 1996a. Female physiology before pregnancy; and the female hormones. *Textbook of Medical Physiology*. 9th ed. Philadelphia: W.B. Saunders, 1003–1016. 1017–1032.
- Guyton Arthur C., John E. Hall. 1996b. Reproductive and Hormonal Functions of the Male (and the Pineal Gland). *Textbook of Medical Physiology*. 9th ed. Philadelphia: W.B. Saunders, 1003–1016.
- Haddy JA, Pankhurst NW. 1999. Stress-induced changes in concentrations of plasma sex steroids in black bream. *J Fish Biol* 55:1304–1316.
- Hagino N. 1976. Effect of thyroid hormone on initiation of persistent oestrus in the rat. *Acta Biol Acad Sci Hung* 27:93–99.
- Hammond GL. 2011. Diverse roles for sex hormone-binding globulin in reproduction. *Biol Reprod* 85:431–441.
- Hapon MB, Gamarra-Luques C, Jahn GA. 2010. Short term hypothyroidism affects ovarian function in the cycling rat. *Reprod Biol Endocrinol* 8:14.
- Hatsuta M, Abe K, Tamura K, Ryuno T, Watanabe G, Taya K, Kogo H. 2004. Effects of hypothyroidism on the estrous cycle and reproductive hormones in mature female rat. *Eur J Pharmacol* 486:343–348.
- Hecker M, Newsted JL, Murphy MB, Higley EB, Jones PD, Wu R, Giesy JP. 2006. Human adrenocarcinoma (H295R) cells for rapid *in vitro* determination of effects on steroidogenesis: hormone production. *Toxicol Appl Pharmacol* 217:114–124.
- Hecker M, Hollert H, Cooper R, Vinggaard AM, Akahori Y, Murphy M, Nellemann C, Higley E, Newstad J, Laskey J, Buckalew A, Grund S, Maletz S, Giesy J, Timm G. 2011. The OECD validation program of the H295R steroidogenesis assay: phase 3. Final inter-laboratory validation study. *Environ Sci Pollut Res Int* 18:503–515.
- Hiipakka RA, Liao S. 1995. Androgenic and anabolic steroids and antagonists. Chapter 50. In *Principles of pharmacology basic concepts & clinical applications*. Munson, PL., Mueller, RA, Breese, GR, Eds. New York, NY: Chapman and Hall.
- Hilscherova K, Jones PD, Gracia T, Newsted JL, Zhang X, Sanderson JT, Yu RM, Wu RS, Giesy JP. 2004. Assessment of the effects of chemicals on the expression of ten steroidogenic genes in the H295R cell line using real-time PCR. *Toxicol Sci* 81:78–89.
- Holson RR, Gough B, Sullivan P, Badger T, Sheehan DM. 1995. Prenatal dexamethasone or stress but not ACTH or corticosterone alter sexual behavior in male rats. *Neurotoxicol Teratol* 17:393–401.
- Hutchinson TH, Ankley GT, Segner H, Tyler CR. 2006. Screening and testing for endocrine disruption in fish—biomarkers as “signposts,” not “traffic lights,” in risk assessment. *Environ Health Perspect* 114:106–114.
- Jahnke GD, Choksi NY, Moore JA, Shelby MD. 2004. Thyroid toxicants: assessing reproductive health effects. *Environ Health Perspect* 112:363–368.
- Jennings DH, Hanken J. 1998. Mechanistic basis of life history evolution in anuran amphibians: thyroid gland development in the direct-developing frog, *Eleutherodactylus coqui*. *Gen Comp Endocrinol* 111:225–232.
- Jensen KM, Kahl MD, Makynen EA, Korte JJ, Leino RL, Butterworth BC, Ankley GT. 2004. Characterization of responses to the antiandrogen flutamide in a short-term reproduction assay with the fathead minnow. *Aquat Toxicol* 70:99–110.
- Juberg DR, Gehen SC, Coady KK, Lebaron MJ, Kramer VJ, Lu H, Marty MS. 2013. Chlorpyrifos: weight of evidence evaluation of potential interaction with the estrogen, androgen, or thyroid pathways. *Regul Toxicol Pharmacol* 66:249–263.
- Kanno J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W. 2001. The OECD program to validate the rat uterotrophic bioassay to screen compounds for *in vivo* estrogenic responses: phase 1. *Environ Health Perspect* 109(8):785–794.
- Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003. The OECD program to validate the rat uterotrophic bioassay. Phase 2: coded single-dose studies. *Environ Health Perspect* 111(12):1550–1558.
- Kicman AT. 2010. Biochemical and physiological aspects of endogenous androgens. *Handb Exp Pharmacol* 195:25–64.
- Kim HS, Shin JH, Moon HJ, Kim TS, Kang IH, Seok JH, Kim IY, Park KL, Han SY. 2002. Evaluation of the 20-day pubertal female assay in Sprague-Dawley rats treated with DES, tamoxifen, testosterone, and flutamide. *Toxicol Sci* 67:52–62.
- Korach KS, Migliaccio S, Davis VL. 1995. Estrogens. In: Munson, PL., Mueller, RA, Breese, GR, Eds. *Principles of Pharmacology: Basic Concepts & Clinical Applications*. New York: Chapman & Hall 809–825.
- Kuiper GG, Carlsson B, Grandien K, Enmark E, Häggblad J, Nilsson S, Gustafsson JA. 1997. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology* 138:863–870.
- Lanzino M, Maris P, Sirianni R, Barone I, Casaburi I, Chimento A, Giordano C, Morelli C, Sisci D, Rizza P, Bonofiglio D, Catalano S, Ando S. 2013. DAX-1, as an androgen-target gene, inhibits aromatase expression: a novel mechanism blocking estrogen-dependent breast cancer cell proliferation. *Cell Death Dis* 4:e724.
- Laws SC, Yavanhaxay S, Copper RL, Eldridge JC. 2006. Nature of the binding interaction for 50 structurally diverse chemicals with rat estrogen receptors. *Toxicol Sci* 94(1):46–56.
- Laws SC, Stoker TE, Ferrell JM, Hotchkiss MG, Cooper RL. 2007. Effects of altered food intake during pubertal development in male and female wistar rats. *Toxicol Sci* 100:194–202.
- Lethimonier C, Flouriot G, Valotaire Y, Kah O, Ducouret B. 2000. Transcriptional interference between glucocorticoid receptor and estradiol receptor mediates the inhibitory effect of cortisol on fish vitellogenesis. *Biol Reprod* 62:1763–1771.

- Li Y, Kumazawa T, Ishiguro T, Kawakami Y, Nishitani H, Tagawa Y, Matsumoto Y. 2011. Hypothyroidism caused by phenobarbital affects patterns of estrous cyclicity in rats. *Congenit Anom (Kyoto)* 51:55–61.
- Lubzens E, Young G, Bobe J, Cerdà J. 2010. Oogenesis in teleosts: how eggs are formed. *Gen Comp Endocrinol* 165:367–389.
- Lutz I, Kloas W, Springer TA, Holden LR, Wolf JC, Krueger HO, Hosmer AJ. 2008. Development, standardization and refinement of procedures for evaluating effects of endocrine active compounds on development and sexual differentiation of *Xenopus laevis*. *Anal Bioanal Chem* 390:2031–2048.
- MacGregor JJ, Jordan VC. 1998. Basic guide to the mechanisms of anti-estrogen action. *Pharmacol Rev* 50:151–196.
- Martinović D, Blake LS, Durhan EJ, Greene KJ, Kahl MD, Jensen KM, Makynen EA, Villeneuve DL, Ankley GT. 2008. Reproductive toxicity of vinclozolin in the fathead minnow: confirming an anti-androgenic mode of action. *Environ Toxicol Chem* 27:478–488.
- Marty MS. 2013. Mammalian methods for detecting and assessing endocrine-active compounds. Endocrine disruptors. Hazard testing and assessment methods. In: Matthiessen P, editor. Hoboken, NJ: John Wiley & Sons, Inc. p 304–340.
- Marty MS, Crissman JW, Carney EW. 1999. Evaluation of the EDSTAC female pubertal assay in CD rats using 17beta-estradiol, steroid biosynthesis inhibitors, and a thyroid inhibitor. *Toxicol Sci* 52:269–277.
- Marty MS, Crissman JW, Carney EW. 2001a. Evaluation of the male pubertal assay's ability to detect thyroid inhibitors and dopaminergic agents. *Toxicol Sci* 60:63–76.
- Marty MS, Crissman JW, Carney EW. 2001b. Evaluation of the male pubertal onset assay to detect testosterone and steroid biosynthesis inhibitors in CD rats. *Toxicol Sci* 60:285–295.
- Marty MS, Johnson KA, Carney EW. 2003. Effect of feed restriction on Hershberger and pubertal male assay endpoints. *Birth Defects Res B Dev Reprod Toxicol* 68:363–374.
- Marty MS, Carney EW, Rowlands JC. 2011. Endocrine disruption: historical perspectives and its impact on the future of toxicology testing. *Toxicol Sci* 120(Suppl 1):S93–S108.
- McCarty LS, Borgert CJ, Mihaich EM. 2012. Information quality in regulatory decision-making: peer-review versus good laboratory practice. *Environ Health Perspect* 120:927–934.
- Milla S, Wang N, Mandiki SNM, Kestemont P. 2009. Corticosteroids: friends or foes of teleost fish reproduction? *Comp Biochem Physiol Part A* 153:242–251.
- Miller WL, Auchus RJ. 2011. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev* 32:81–151.
- Mills LJ, Chichester C. 2008. Review of evidence: are endocrine-disrupting chemicals in the aquatic environment impacting fish populations? *Sci Total Environ* 343:1–34.
- Negus SS. 2006. Some implications of receptor theory for in vivo assessment of agonists, antagonists and inverse agonists. *Biochem Pharmacol* 71:1663–1670.
- O'Connor JC, Davis LG, Frame SR, Cook JC. 2000. Evaluation of a tier I screening battery for detecting endocrine-active compounds (EACs) using the positive controls testosterone, coumestrol, progesterone, and RU486. *Toxicol Sci* 54:338–354.
- Odum J, Lefevre PA, Tittensor S, Paton D, Routledge EJ, Beresford NA, Sumpter JP, Ashby J. 1997. The rodent uterotrophic assay: critical protocol features, studies with nonyl phenols, and comparison with a yeast estrogenicity assay. *Regul Toxicol Pharmacol* 25:176–188.
- Organization for Economic Cooperation and Development (OECD). 2004a. Detailed review paper on fish screening assays for the detection of endocrine active substances. Available at: <http://www.oecd-ilibrary.org/docserver/download/9750471e.pdf?expires=1382988166&id=id&acname=guest&checksum=0797CBA2C275400AB5EFB3BE62C5B5F0>. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2004b. Detailed review paper on amphibian metamorphosis assay for the detection of thyroid active substances. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/edmvac/oecd_amphibian_drp.pdf. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2007a. Final report of the validation of the 21-day fish screening assay for the detection of endocrine active substances. Series on testing and assessment number 78. ENV/JM/MONO(2007)25.
- Organization for Economic Cooperation and Development (OECD). 2007b. OECD guideline for the testing of chemicals. Uterotrophic bioassay in rodents: a short-term screening test for oestrogenic properties. Available at: <http://www.oecd-ilibrary.org/docserver/download/9744001e.pdf?expires=1382988559&id=id&acname=guest&checksum=701B965B460ABEA80055190B33E45266>. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2007c. Validation of the amphibian metamorphosis assay as a screen for thyroid-active chemicals: integrated summary report. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/ama_isr.pdf. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2009a. OECD guideline for the testing of chemicals Hershberger bioassay in rats: a short-term screening assay for (anti)androgenic properties. Available at: <http://www.oecd-ilibrary.org/docserver/download/9744101e.pdf?expires=1382988624&id=id&acname=guest&checksum=F2F2873C27F2B066BCF29AC1D3179C1A>. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2009b. OECD guideline for the testing of chemicals 21-day fish assay: a short-term screening for oestrogenic and androgenic activity, and aromatase inhibitions. Available at: <http://www.oecd-ilibrary.org/docserver/download/9723001e.pdf?expires=1382988737&id=id&acname=guest&checksum=70DAFD1299CE50ABB86606FD2102B80>. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2009c. OECD guideline for the testing of chemicals the amphibian metamorphosis assay. Available at: <http://www.oecd-ilibrary.org/docserver/download/9723101e.pdf?expires=1382988799&id=id&acname=guest&checksum=842FFDCB326C1A944186556C5FDA5823>. Accessed October 28, 2013.
- Organization for Economic Cooperation and Development (OECD). 2012. Guidance document on standardised test guidelines for evaluating chemicals for endocrine disruption no 150. Organisation for Economic Cooperation and Development, Paris, 24-Aug-2012. Available at: [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=ENV/JM/MONO\(2012\)22&doclanguage=en](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=ENV/JM/MONO(2012)22&doclanguage=en). Accessed October 28, 2013.
- Ososki AL, Kennelly EJ. 2003. Phytoestrogens: a review of the present state of research. *Phytother Res* 17:845–869.
- Owens W, Ashby J. 2002. Critical review and evaluation of the uterotrophic bioassay for the identification of possible estrogen agonists and antagonists: in support of the validation of the OECD uterotrophic protocols for the laboratory rodent. Organisation for Economic Co-operation and Development. *Crit Rev Toxicol* 32(6): 445–520.
- Owens W, Ashby J, Odum J, Onyon L. 2003. The OECD program to validate the rat uterotrophic bioassay. Phase 2: dietary phytoestrogen analyses. *Environ Health Perspect* 111(12):1559–1567.
- Pankhurst NW, Van Der Kraak G. 2000. Evidence that acute stress inhibits ovarian steroidogenesis in rainbow trout *in vivo*, through action of cortisol. *Gen Comp Endocrinol* 117:225–237.
- Parks LG, Cheek AO, Denslow ND, Heppell SA, McLachlan JA, LeBlanc GA, Sullivan CV. 1999. Fathead minnow (*Pimephales promelas*) vitellogenin: purification, characterization and quantitative immunoassay for the detection of estrogenic compounds. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 123:113–125.
- Pereyra-Martinez AC, Roselli CE, Stadelman HL, Resko JA. 2001. Cytochrome P450 aromatase in testis and epididymis of male rhesus monkeys. *Endocrine* 16:15–19.
- Pickford DB. 2010. Screening chemicals for thyroid-disrupting activity: a critical comparison of mammalian and amphibian models. *Crit Rev Toxicol* 40:845–892.
- Quignot N, Desmots S, Barouki R, Lemazurier E. 2012. A comparison of two human cell lines and two rat gonadal cell primary cultures as *in vitro* screening tools for aromatase modulation. *Toxicol In Vitro* 26:107–118.
- Reel JR, Lamb IV JC, Neal BH. 1996. Survey and assessment of mammalian estrogen biological assays for hazard characterization. *Fundam Appl Toxicol* 34:288–305.
- Rocca MS, Pepperl S. 2000. Final report assessment of pubertal development and thyroid function in juvenile male rats. Sponsored by Environmental Protection Agency RTP: MD-71 NHEERL Research Triangle Park, NC 27711. TherImmune Research Corporation, Study Number 1143–100, 15 Firstfield Road, Gaithersburg, MD 20878.
- Rozman KK, Bhatia J, Calafat AM, Chambers C, Culty M, Etzel RA, Flaws JA, Hansen DK, Hoyer PB, Jeffery EH, Kesner JS, Marty S, Thomas JA and Umbach D. 2006. NTP-CERHR expert panel report on the reproductive and developmental toxicity of genistein. *Birth Defects Res B Dev Reprod Toxicol* 77:485–638.

- Schmid T, Gonzalez-Valero J, Rufli H, Dietrich DR. 2002. Determination of vitellogenin kinetics in male fathead minnows (*Pimephales promelas*). *Toxicol Lett* 131:65–74.
- Schmidt WN, Katzenellenbogen BS. 1979. Androgen-uterine interactions: an assessment of androgen interaction with the testosterone- and estrogen-receptor systems and stimulation of uterine growth and progesterone-receptor synthesis. *Mol Cell Endocrinol* 15:91–108.
- Shappell NW, Hyndman KM, Bartell SE, Schoenfuss HL. 2010. Comparative biological effects and potency of 17 α - and 17 β -estradiol in fathead minnows. *Aquat Toxicol* 100:1–8.
- Shayu D, Rao AJ. 2006. Expression of functional aromatase in the epididymis: role of androgens and LH in modulation of expression and activity. *Mol Cell Endocrinol* 249:40–50.
- Shin JH, Kim HS, Moon HJ, Kang H, Kim TS, Seok JH, Kim IJ, Park KL, Han SY, Nam SY. 2002. Effects of flutamide on puberty in male rats: an evaluation of the protocol for the assessment of pubertal development and thyroid function. *J Toxicol Environ Health A* 65:433–445.
- Stoker T, Zorrilla L. 2010. The effects of endocrine disrupting chemicals on pubertal development in the rat. *Endocrine toxicology*. 3rd ed. Eldridge, C., Steven, JT., Eds. CRC Press. p 27–81.
- Stoker TE, Parks LG, Gray LE, Cooper RL. 2000. Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC recommendations. *Endocrine Disrupter Screening and Testing Advisory Committee. Crit Rev Toxicol* 30:197–252.
- Stromborg KL. 1986. Reproduction of bobwhites fed different dietary concentrations of an organophosphate insecticide, methamidophos. *Arch Environ Contam Toxicol* 15:143–147.
- Tata JR. 2006. Amphibian metamorphosis as a model for the developmental actions of thyroid hormone. *Mol Cell Endocrinol* 246:10–20.
- U.S. EPA. 1999. Review of the EPA's proposed environmental disruptor screening program; review of the Endocrine Disruptor Screening Program by a joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. Available at: [http://yosemite.epa.gov/sab/sabproduct.nsf/C8ABD410E357D-BCF85257193004C42C4/\\$File/ec13.pdf](http://yosemite.epa.gov/sab/sabproduct.nsf/C8ABD410E357D-BCF85257193004C42C4/$File/ec13.pdf). Accessed July 18, 2010.
- U.S. EPA. 2006. Fish screening assay discussion paper.
- U.S. EPA. 2007a. Endocrine Disruptor Screening Program. Validation of the fish short term reproduction assay: integrated summary report. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/fish_assay_isr.pdf. Accessed October 28, 2013.
- U.S. EPA. 2007b. Integrated summary report for validation of a test method for assessment of pubertal development and thyroid function in juvenile male rats as a potential screen in the Endocrine Disruptor Screening Program Tier-1 Battery. Office of Science Coordination and Policy and Office of Research and Development. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/male_pubertal_isr.pdf. Accessed October 28, 2013.
- U.S. EPA. 2007c. Integrated summary report for validation of a test method for assessment of pubertal development and thyroid function in juvenile female rats as a potential screen in the Endocrine Disruptor Screening Program Tier-1 Battery. Office of Science Coordination and Policy and Office of Research and Development. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/female_isr_v4.1c.pdf. Accessed October 28, 2013.
- U.S. EPA. 2009. Series 890-Endocrine Disruptor Screening Test OPPTS 890.1500: pubertal development and thyroid function in intact juvenile/peripubertal male rats. Available at: <http://www.regulations.gov/#/documentDetail;D=EPA-HQ-OPPT-2009-0576-0010>. Accessed October 28, 2013.
- U.S. EPA (Environmental Protection Agency). 2011a. Weight-of-evidence: evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. EPA-HQ-OPPT-2010-0877-0021. Available at: <http://www.regulations.gov/#/documentDetail;D=EPA-HQ-OPPT-2010-0877-0021>. Accessed October 22, 2013.
- U.S. EPA (Environmental Protection Agency). 2011b. Endocrine Disruptor Screening Program Test Guidelines. OCSPP 890.1400: Hershberger assay. Standard Evaluation Procedure. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/toresources/seps/Final.890.1600_Hershberger_Assay_SEP.10.6.11.pdf. Accessed October 28, 2013.
- U.S. EPA (Environmental Protection Agency). 2011c. Endocrine Disruptor Screening Program Test Guidelines. OCSPP 890.1150: androgen receptor binding assay (rat ventral prostate cytosol. Standard Evaluation Procedure. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/toresources/seps/Final.890.1150_AR.Binding_Assay_SEP.10.5.11.pdf. Accessed October 28, 2013.
- U.S. National Library of Medicine. 2013. MeSH Headings—NCBI: “Androgen Antagonists” definition. Available at: <http://www.ncbi.nlm.nih.gov/mesh/?term=androgen+antagonists>. Accessed October 28, 2013.
- Vitale DC, Piazza C, Melilli B, Drago F, Salomone S. 2012. Isoflavones: estrogenic activity, biological effect and bioavailability. *Eur J Drug Metab Pharmacokinet* 38:15–25.
- Wade GN, Blaustein JD, Gray JM, Meredith JM. 1993. ICI 162,780: a pure antiestrogen that affects behaviors and energy balance in rats without acting in the brain. *Am J Physiol* 265:R1392–R1398.
- Wang S, Aarts JM, Evers NM, Peijnenburg AA, Rietjens IM, Bovee TF. 2012. Proliferation assays for estrogenicity testing with high predictive value for the *in vivo* uterotrophic effect. *J Steroid Biochem Mol Biol* 128:98–106.
- Wason S, Pohlmeier-Esch G, Pallen C, Palazzi X, Espuña G, Bars R. 2003. 17 α -methyltestosterone: 28-day oral toxicity study in the rat based on the “Enhanced OECD Test Guideline 407” to detect endocrine effects. *Toxicology* 192:119–137.
- Wilbur HM, Collins JP. 1973. Ecological aspects of amphibian metamorphosis: nonnormal distributions of competitive ability reflect selection for facultative metamorphosis. *Science* 182:1305–1314.
- Winters SJ. 1998. Androgens and antiandrogens chapter 39. *Human pharmacology: molecular to clinical*. In: Brody TM, editor. St. Louis, MO: Mosby-Year Book, Inc.
- Wu RSS, Zhou BS, Randall DJ, Woo NTY, Lam PKS. 2003. Aquatic hypoxia is an disrupter and impairs fish reproduction. *Environ Sci Technol* 37:1137–1141.
- Yen PM. 2001. Physiological and molecular basis of thyroid hormone action. *Physiol Rev* 81:1097–1142.
- York RG, Brown WR, Girard MF, Dollarhide JS. 2001. Two-generation reproduction study of ammonium perchlorate in drinking water in rats evaluates thyroid toxicity. *Int J Toxicol* 20:183–197.