

Original Article

Key Lessons from Performance of the U.S. EPA Endocrine Disruptor Screening Program (EDSP) Tier 1 Male and Female Pubertal Assays

Donald G. Stump,¹ John C. O'Connor,² Joseph M. Lewis,² and M. Sue Marty^{3*}

¹Developmental and Reproductive Toxicology, WIL Research Laboratories, LLC, Ashland, Ohio

²DuPont Haskell Global Centers for Health and Environmental Sciences, Newark, Delaware

³Toxicology & Environmental Research and Consulting, The Dow Chemical Company, Midland, Michigan

The male and female pubertal assays, which are included in the U.S. Environmental Protection Agency's (EPA) Endocrine Disruptor Screening Program (EDSP) Tier 1 battery, can detect endocrine-active compounds operating by various modes of action. This article uses the collective experience of three laboratories to provide information on pubertal assay conduct, interlaboratory reproducibility, endpoint redundancy, and data interpretation. The various criteria used to select the maximum tolerated dose are described. A comparison of historical control data across laboratories confirmed reasonably good interlaboratory reproducibility. With a reliance on apical endpoints, interpretation of pubertal assay effects as specifically endocrine-mediated or secondary to other systemic effects can be problematic and mode of action may be difficult to discern. Across 21–23 data sets, relative liver weight, a nonspecific endocrine endpoint, was the most commonly affected endpoint in male and female assays. For endocrine endpoints, patterns of effects were generally seen; rarely was an endocrine-sensitive endpoint affected in isolation. In males, most frequently missed EPA-established performance criteria included mean weights for kidney and thyroid, and the coefficient of variation for age and body weight at prepubertal separation, seminal vesicle weight, and final body weight. In females, the frequently missed EPA-established performance criteria included mean adrenal weight and mean age at vaginal opening. To ensure specificity for endocrine effects, the pubertal assays should be interpreted using a weight-of-evidence approach as part of the entire EDSP battery. Based on the frequency with which certain performance criteria were missed, an EPA review of these criteria is warranted. *Birth Defects Res (Part B)* 101:43–62, 2014. © 2014 The Authors. *Birth Defects Research (Part B)* published by Wiley Periodicals, Inc.

Key words: *male pubertal assay; female pubertal assay; endocrine disruptor; endocrine screening; EDSP; estrogen; androgen; thyroid*

INTRODUCTION

Responding to concerns that man-made chemicals found in the environment may have the potential to impact endocrine function in humans and/or wildlife, Congress mandated that the United States Environmental Protection Agency (U.S. EPA) develop an Endocrine Disruptor Screening Program (EDSP). The EDSP, which was launched in 2009, consists of two tiers of assays and tests to examine potential endocrine activity.

Tier 1 is composed of a screening battery of 11 *in vitro* and *in vivo* assays collectively designed to identify substances that have potential to interact with components of the estrogen, androgen, and thyroid hormone signaling pathways (U.S. EPA, 2011a). Results of EDSP Tier 1, along with other scientifically relevant information that may already be available for a particular compound, are used in a weight-of-evidence (WoE) determination of a substance's potential to interact with these systems, and will

be used to trigger specific Tier 2 tests if warranted by experimental findings. EDSP Tier 2 evaluates dose–response relationships and identifies adverse effects in studies of longer duration and increased complexity, which will form the basis for risk assessments on these compounds (U.S. EPA, 2011a).

The male and female pubertal assays are included as part of the Tier 1 battery and are designed to detect endocrine-active compounds that operate through a variety of modes of action (MoAs), including potential estrogenic/antiestrogenic effects (primarily the female assay),

*Correspondence to: M. Sue Marty, Toxicology & Environmental Research and Consulting, The Dow Chemical Company, 1803 Building, Midland, Michigan 48674. E-mail: mmarty@dow.com

Received 27 October 2013; Accepted 02 January 2014

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/bdrb) DOI: 10.1002/bdrb.21097

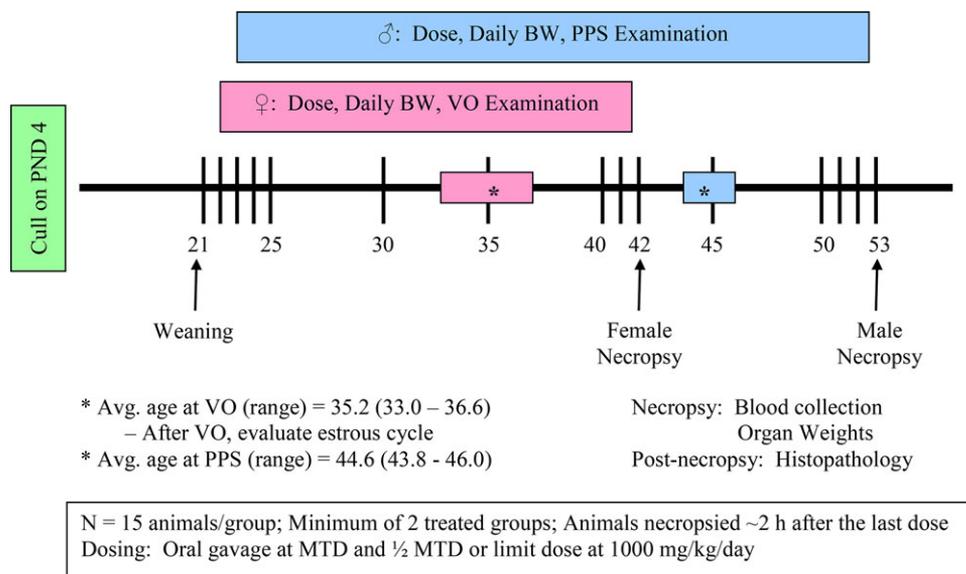


Fig. 1. Study design for the male and female pubertal assays. BW, body weight; VO, vaginal opening; PPS, preputial separation.

androgen/antiandrogen effects (primarily the male assay), modulation of steroid biosynthesis, alterations in the hypothalamic–pituitary–gonadal axis, and thyroid perturbations.

To conduct the male and female pubertal assays (Fig. 1), male or female weanling rats are randomly assigned to treatment groups in a manner that yields similar mean body weights and variances across groups; littermates are not assigned to the same dose group. Rats are exposed to the test compound by oral gavage from post-natal day (PND) 23–53 (males) or 22–42 (females). Beginning on PND 30 (males) or PND 22 (females), animals are evaluated daily for puberty onset, which is indicated by preputial separation (PPS) in the males and vaginal opening in the females. When puberty onset is achieved, the animal's age and body weight are recorded. Once vaginal opening is complete, daily vaginal smears are collected to monitor age at first estrus and to evaluate the pattern and regularity of the estrous cycle. Males and females are necropsied on PND 53 and 42, respectively. A terminal blood sample is collected for clinical chemistry and serum hormone analyses (thyroid-stimulating hormone (TSH) and thyroxine (T4) in both the males and females and testosterone in the males). The liver, kidneys, adrenals, pituitary, and thyroid are weighed in both sexes. Other organ weights include ovaries and uterus (with and without fluid) in the females, and testes, epididymides, ventral prostate, dorsolateral prostate, seminal vesicles with coagulating glands (with and without fluid), and levator ani-bulbocavernosus (LABC) muscles in the males. Tissues examined histopathologically include the ovary, uterus, kidney, and thyroid for the females, and the testis, epididymis, kidney, and thyroid for the males. Test guidelines are available that describe the conduct, interpretation, and performance specifications for these assays (U.S. EPA, 2009a, 2009b).

The purpose of the current report is to evaluate the EDSP Tier 1 male and female pubertal assays based on the authors' experiences in fulfilling EPA-mandated en-

docrine screening under EDSP. The initial EDSP prioritization list for screening (EDSP List 1) was announced in 2009 with a list of 67 compounds, and test orders were subsequently issued starting in October 2009. Registrants or manufacturers provided EDSP Tier 1 screening data for 52 compounds. The current report presents experiences from three laboratories, which have conducted approximately 40% of the EDSP List 1 test orders. The objective of this article is to provide information on assay conduct based on the collective experiences of three laboratories, data on interlaboratory reproducibility and endpoint redundancy, and factors to consider when interpreting male and female pubertal assay data.

Assay Conduct: Implementation

Before initiating the male and female pubertal assays, some capability development may be needed within the test facility. While many of the endpoints are included in other study types, hormone measurements (T4, TSH, and testosterone) are new regulatory requirements that have been included in the pubertal assays. Laboratories may choose to develop these methods in-house or serum samples can be sent to a contract research organization for hormone measurements. Laboratories also are required to develop and use a five-point thyroid histopathology scoring system, which uses a graded scale for both follicular cell height and colloid amount (U.S. EPA 2009a, 2009b).

Assay Conduct: Experimental Conditions

Specific conditions for animal husbandry as required by the test guidelines for the male and female pubertal assays are unique to the new endocrine test guidelines. While low phytoestrogen diet is not specifically required for the pubertal assays, the test guidelines require that the genistein-equivalent content of the diet must be less than or equal to 300 ppm. Many standard laboratory diets exceed the 300 ppm genistein-equivalent limit, at least in some feed lots. If a "low phytoestrogen" diet is not used

for the pubertal assays, laboratories are encouraged to analyze each lot of diet for phytoestrogen content to ensure compliance with the 300 ppm genistein-equivalent requirement. In addition, tap water is not acceptable for the pubertal assays according to the test guidelines. Deionized drinking water is required and must not be administered using polycarbonate water supply equipment due to concerns about variability in water quality (e.g., possible presence of disinfection byproducts, perchlorate, etc.) that were expressed by the Endocrine Disruptor Methods Validation Subcommittee (U.S. EPA, 2011b).

The test guidelines state that corn cob bedding should not be used due to its potential to disrupt endocrine activity (Markaverich et al., 2002). Instead, heat-treated laboratory-grade wood shavings other than cedar are the recommended bedding. Cedar bedding has been reported to result in high rat pup mortality (Burkhart and Robinson, 1978). One of the authors of this article had similar issues with high pup mortality when using heat-treated pine shavings as bedding for two pubertal assays. For one of these assays, pup survival was low enough that the start of the study needed to be delayed to purchase additional time-mated animals. However, when hardwood bedding (Aspen bedding) was used, pup survival was high. Therefore, in addition to the bedding limitations specified in the test guideline, the authors recommend avoiding the use of pine bedding. The test guidelines also advocate the use of clear plastic containers for animal housing, but suitable housing that meets the criteria specified in the *Guidelines for the Care and Use of Laboratory Animals* for co-housing animals of that age/size is appropriate.

Some methodological changes to the test guidelines have been approved by the U.S. EPA (2011b), including the use of 12-hr light:12-hr dark cycle (instead of the 14-hr light:10-hr dark cycle given in the test guidelines) and the option to use the most appropriate gavage needle size for the size of animal being dosed (vs. the specification in the EPA test guideline of an 18 gauge gavage needle, 1 to 1½ inch length with a 2.25 mm ball). Laboratories also are advised to review test guideline requirements for anesthesia and necropsy procedures. The U.S. EPA has approved the use of isoflurane for necropsy anesthesia and other forms of blood collection aside from decapitation are acceptable (e.g., aortal exsanguination or other methods that yields sufficient blood volume). Fixed pituitary weights, rather than fresh weights, may be reported.

Assay Conduct: Use of Litters for Animal Selection for the Pubertal Assays

The pubertal assay test guidelines require that test animals must be born in-house to avoid shipping stress during late gestation and lactation. Most laboratories performing the pubertal assay will not breed females in-house but will instead use females bred at the animal supplier (time-mated). According to the test guidelines, if time-mated females are used, then all dams should arrive on the same gestation day. The guidelines also require the offspring to be necropsied on PND 42 (females) or PND 53 (males) between 0900 and 1300 hr. To comply with the time of necropsy, laboratories are permitted to schedule necropsies so that animals are necropsied over 2 days (i.e.,

one half of the females are necropsied on PND 42 and one half of the females are necropsied on PND 43 (or PND 53 and 54 for the males)). A practical alternative approach is to divide the animals based upon the pup's date of birth as a group of dams with the same day of breeding will not deliver all offspring on the same day. Thus, pups born on either gestation day 21 or 22 could be used for study, with all animals necropsied on either PND 42 (females) or PND 53 (males); necropsies will be divided over 2 days based on the pups' dates of birth. This latter approach also may be more manageable for laboratories, given the time limits for sample collection and the number of endpoints measured in these assays.

The current design requiring the use of pregnant females increases the number of animals needed for pubertal assays. For example, using the current design (control and two dose groups; 15 animals/group = 45 animals with no littermates assigned to the same dose group), then a minimum of 15 litters are needed (assuming three pups/sex/litter, one pup can be assigned to each dose group in a male and female pubertal assay). Assuming an average of 14 pups/litter (CD rats), this could total 162 animals that are not used in the study (18 adults +8 pups/litter × 18 litters, if three extra dams are ordered to allow for differences in delivery dates and any nonpregnant females). Given the study design requirements, the optimal use of animals is achieved when a male and female pubertal assay can be conducted concurrently. If the pubertal assays must be conducted separately, the number of animals not used in the study could increase to 207. Thus, careful planning is warranted to ensure the successful conduct of each pubertal assay when it is initiated to minimize animal usage.

To improve animal use, consider whether weanling rats not assigned to pubertal assays could be used for dose range-finding studies for other compounds on the priority list for endocrine screening. Alternatively, extra female rats could be assigned to a uterotrophic assay if dosing is initiated no later than PND 22 (per the Organisation for Economic Co-Operation and Development (OECD) test guideline 440, dosing must be completed before PND 25; OECD, 2007). A further option is to consider using an additional dose group (three test compound groups plus a control group) for the pubertal assay. While not required by the guidelines, the additional dose group can be helpful in identifying a maximum tolerated dose (MTD).

Assay Conduct: Dose Selection and Number of Dose Groups

Dose selection is an aspect of the pubertal assays that can be challenging. The test guideline specifies that the high-dose level should be set at or just below the MTD, and provides the following guidance:

- (1) The high dose does not exceed the limit dose of 1 g/kg/day.
- (2) The high-dose level causes a statistically significant reduction in terminal body weight gain, the reduction is no greater than approximately 10% of the mean terminal body weight for the controls, and there are no associated clinical signs of toxicity (although the male pubertal assay test guideline also states that a

Table 1
Dose-Setting Criteria for Male Rats

Compound ^a	Range-finder performed	MTD achieved (no, yes, exceeded)	Endpoint that was used to select dose levels	No. of dose levels ^b
1	Yes	Yes	Body weight	3
2	Yes	No	Kidney pathology	3
3	Yes	No	Clinical signs	3
4	Yes	Yes	Body weight	3
5	No	Yes	Cholinesterase inhibition	4
6	Yes	Yes	Body weight	4
7	Yes	Yes	Liver weight	3
8	No	No	Clinical signs	3
9	Yes	Yes ^c	Clinical signs	3
10	No	Yes ^d	Limit dose (1000 mg/kg)	3
14	Yes	Yes	Clinical signs/body weight	4
15	Yes	Yes	Creatinine increases	4
16	Yes	Yes	Cholinesterase inhibition	4
17	Yes	Yes	Hepatocellular necrosis	4
18	Yes	Yes	Body weight	4
19	Yes	Yes	Body weight	4
20 (assay #1)	Yes	Exceeded	Body weight	4
20 (assay #2)	Yes	Yes	Body weight	4
21	No	Yes ^e	Cholinesterase inhibition	3
22	No	Yes ^e	Cholinesterase inhibition	3
23	Yes	Yes	Clinical signs	3

^aTest materials were arbitrarily assigned a "compound number," which is consistent between this table and Table 2.

^bNumber of dose levels evaluated including controls in the male pubertal assay.

^cDose levels selected to avoid systemic toxicity that was observed at slightly higher dose levels in previous studies.

^dMaximum tolerated dose was achieved since testing was done at limit dose of 1000 mg/kg.

^eMaximum tolerated dose selected using read across to previously performed studies.

decrease in terminal body weight gain of approximately 6% may require additional information/data for assay interpretation).

- (3) The MTD may be exceeded if abnormal blood chemistry values are seen at termination (particularly creatinine and blood urea nitrogen (BUN)).
- (4) The MTD may be exceeded if histopathology of the kidney (or any other organ where gross observations indicate damage) is seen.

The second dose level is set to one half of the high-dose (MTD) level.

In practice, laboratories relied on numerous criteria when selecting dose levels for the pubertal assays, including liver weight increases, liver histopathology, and cholinesterase inhibition (Tables 1 and 2). For the studies performed by the authors, the MTD criteria were usually the same for the male and female pubertal assays for any given test compound; however, there were gender differences in the determining factor for the MTD for two test compounds (compounds 2 and 6). Aside from the parameters identified in the test guidelines, previous toxicity data on a test compound, possibly coupled with the inclusion of additional endpoints in the range-finding study or pubertal assays, can aid in identification of an appropriate MTD criterion to avoid significant systemic toxicity that can contribute to nonspecific alterations in endocrine-sensitive endpoints. For example, read-across from previously performed studies were used to select doses for two compounds (21 and 22), and an MTD was successfully achieved in the pubertal studies.

Although the EPA pubertal assay test guidelines do not require range-finding studies, for approximately 75% of the test compounds, laboratories conducted dose range-finding studies to select dose levels for the pubertal assays (Tables 1 and 2; range-finding studies for 15 of 20 compounds in the male pubertal assay and 18 of 23 compounds in the female pubertal assay). One of the difficulties often encountered in setting dose levels was that the route of exposure from previous studies with the same compound were different from the required administration route (oral gavage) for the EDSP test guidelines. If data were available from previous studies that utilized oral gavage compound administration, these data were typically in adult (postpubertal) animals. As a result of the different toxicokinetics associated with a bolus dose versus a dose administered by dietary consumption, as well as potential differences in sensitivity of immature compared to adult animals, the task of setting dose levels for the pubertal assays can be difficult. Thus, laboratories usually performed range-finding studies, using gavage dosing and animals of a similar age to those used in the main study, to increase confidence that the high-dose level will be at or near the required MTD. With multiple parameters under consideration, the need to achieve, but not exceed, an MTD places a considerable burden on range-finding studies. When range-finding studies identified dose levels that exceeded the MTD, the pubertal assays were conducted at lower dose levels, which sometimes did not meet the MTD criteria in the definitive study (2 of 21 studies in the male pubertal assay, 1 of 23 studies in the female pubertal assay). Conversely, in one male

Table 2
Dose-Setting Criteria for Female Rats

Compound ^a	Range-finder performed	MTD achieved (no, yes, exceeded)	Endpoint that was used to select dose levels	No. of dose levels ^b
1	Yes	Yes	Body weight	3
2	Yes	Yes	Body weight	3
3	Yes	No	Clinical signs	3
4	Yes	Yes	Body weight	3
5	No	Yes	Cholinesterase inhibition	4
6	Yes	Yes ^c	Limit dose (1000 mg/kg)	4
7	Yes	Yes	Liver weight	3
8	No	No	Clinical signs	3
9	Yes	Yes ^d	Clinical signs	3
10	No	Yes ^d	Limit dose (1000 mg/kg)	3
11	Yes	Yes ^d	Body weight	4
12	Yes	Yes	Body weight	5
13	Yes	Yes	Body weight	3
14	Yes	Yes ^d	Clinical signs/body weight	4
15	Yes	Yes	Creatinine increases	4
16	Yes	Yes	Cholinesterase inhibition	4
17	Yes	Yes ^d	Hepatocellular necrosis	4
18	Yes	Yes ^d	Body weight	4
19	Yes	Yes	Body weight	4
20	Yes	Yes	Body weight	4
21	No	Yes ^e	Cholinesterase inhibition	3
22	No	Yes ^e	Cholinesterase inhibition	3
23	Yes	Yes	Clinical signs	3

^aTest materials were arbitrarily assigned a “compound number,” which is consistent between Table 1 and this table.

^bNumber of dose levels evaluated including controls in the female pubertal assay.

^cMaximum tolerated dose was achieved since testing was done at limit dose of 1000 mg/kg.

^dDose levels selected to avoid systemic toxicity that was observed at slightly higher dose levels in previous studies.

^eMaximum tolerated dose selected using read across to previously performed studies.

pubertal assay, the MTD was exceeded despite conducting a range-finding study, which indicates some variability in the responsiveness of juvenile animals across pubertal assays (although it is unclear whether this variability was related to the assay, the test compound or both). While range-finding studies may represent the most predictive approach for determining dose levels for the main study, it does add cost, animal use, and additional time required to perform the pubertal assays.

In a recent endocrine workshop (Juberg et al., 2013), laboratories acknowledged numerous difficulties encountered in dose setting for the pubertal assays. With only two dose levels required by the test guideline, the ramifications of over- or underestimating dose concentrations could result in an invalid study. Several studies we conducted employed three or four test substance dose levels. When three dose levels (control plus two test substance groups) were employed for the male pubertal assays, 3 of 11 studies that had three dose levels did not achieve an MTD, while all 10 studies achieved an MTD when four dose levels (control plus three test substance groups) were used (Table 1). Similarly, for the female pubertal assay, 2 of 12 studies that only had three dose levels (control plus two test substance groups) did not achieve an MTD, while when four or five dose levels (control plus three or four test substance groups) were selected, all 11 studies achieved an MTD (Table 2). In addition, the inclusion of

an additional dose level allows for a better evaluation of dose–response relationships and ensures there are sufficient groups below the MTD to allow interpretation of assay results (i.e., if the high dose exceeded the MTD in a two-dose level study, there is only one dose level from which data can be evaluated for potential endocrine activity). While additional dose groups may aid in data interpretation, registrants are cautioned that the addition of extra dose levels (e.g., three dose levels and a control) makes pubertal assay data applicable or useable for risk assessment purposes if apical effects are judged to be adverse by the U.S. EPA (2013).

Assay Conduct: Ovarian Histopathology

In the female pubertal assay test guideline, the ovarian histopathology text states that “Five random sections (of ovary) are evaluated using the method of Smith et al.” The methodology described by Smith et al. (1991) compared ovarian follicle counts collected using different sectioning/sampling procedures, which raised a question as to whether ovarian follicle counts were required as part of the female pubertal assay. Subsequently, there was a clarification by the EPA (U.S. EPA, 2011b); ovarian follicle counts are not part of the required endpoints under the female pubertal test guideline. The Smith et al. (1991) paper was referenced to support the sampling methodology:

the use of appropriately prepared random sections as opposed to serial sections for ovarian histopathology.

Assay Conduct: Statistics

Age and body weight at puberty onset, and all organ weights are analyzed by analysis of variance (ANOVA) and analysis of covariance (ANCOVA) using body weight at PND 21 as the covariate. Because animals are randomized into dose groups based on body weight (while controlling for litter), adjusting terminal organ weights for PND 21 body weight makes almost no correction for body-weight-mediated changes in absolute organ weights. While the selection of a covariate that is not influenced by treatment (e.g., body weight at weaning) is a customary statistical practice, this analysis does not account for the impact of terminal body weight changes on organ weight endpoints in the pubertal assays (e.g., Laws et al., 2007). Furthermore, the pubertal assay test guidelines require that data such as age at puberty onset and organ weights are analyzed three ways: ANOVA, ANCOVA, and linear trend analyses. It is unclear from the EPA pubertal test guidelines how researchers are to interpret results that are significant using one statistical method and not significant by an alternate analysis. The Scientific Advisory Panel in 2008 recommended that the EPA design new statistics for the pubertal assays that considered body-weight-mediated effects on organ weights. This recommendation could greatly improve the interpretability of assay data. However, currently, EPA has not developed new guidance on statistical analyses for the pubertal assays.

The rationale for not using terminal body weight when analyzing pubertal assay data is that endocrine-active compounds may affect overall body weight gain and terminal body weight. Further discussion is warranted to determine which endocrine (estrogen, androgen, thyroid) MoAs affect growth and whether these MoAs can be identified with other Tier I screening assays, or in the pubertal assay despite body weight changes. For example, environmental estrogens, antiandrogens, and thyroid-active compounds can decrease the rate of growth; however, the effects of these compounds would not be mistaken for systemic toxicity when additional data from the Tier 1 screening assays are evaluated. For example, estrogenic compounds typically accelerate age at vaginal opening, which occurs at a lower body weight (e.g., methoxychlor, ethynyl estradiol, U.S. EPA, 2007a). In addition, results of the uterotrophic assay would support the interpretation of estrogenicity. Antiandrogens would be identified in the Hershberger assay, which is designed to be sensitive to this MoA (O'Connor et al., 1999a). Thyroid-active agents also would decrease T3 and T4, increase TSH, increase thyroid weights, and produce characteristic changes in thyroid histopathology (e.g., propylthiouracil, DE-71; U.S. EPA, 2007a, 2007b). Thus, while these endocrine MoAs can affect growth rate, interpretation of assay results would not be confused with systemic toxicity. Careful consideration should be given to which MoAs decrease growth rate, whether systemic toxicity versus endocrine MoAs can be differentiated, and the potential for false positives if body-weight-mediated changes are not considered during pubertal assay conduct.

ASSAY INTERPRETATION

The male and female pubertal assays underwent a validation program coordinated by the U.S. EPA (2007a, 2007b). For endocrine-active compounds used in the validation program, multiple related endpoints were typically affected and patterns of effects were reproducible across laboratories. However, many challenges have been identified that may impact interpretation of pubertal assay data (Borgert et al., 2011a). These challenges include the reliance of the pubertal assays on apical endpoints, which may make it difficult to identify the specific endocrine MoA or whether effects were due to a primary endocrine MoA or secondary to systemic toxicity. The pubertal assays examine endocrine endpoints in young animals during a dynamic period when integrated function of the endocrine system is required; however, the dynamic nature of growth and hormone levels during this period may result in differences due to slight delays in development. Furthermore, for some endpoints, biological variability in pubertal data may be greater due to this dynamic developmental period. The following sections are designed to present information on factors to consider when interpreting pubertal assay data.

Assay Interpretation: Historical Control Data (HCD)

HCD can be useful when interpreting pubertal assay results, particularly for more variable endpoints (e.g., ventral prostate weights, serum hormone levels, etc.). For two laboratories participating in this publication, interlaboratory HCD are presented in Table 3 (males) and Table 4 (females). Generally, control values for assay endpoints were similar across laboratories for both the male and female pubertal assays, showing reasonably good interlaboratory reproducibility. In the male pubertal assay, the most variable endpoints included dorsolateral prostate weight, weight of the seminal vesicles with coagulating glands (with and without fluid), thyroid weight, creatinine, and serum concentrations of TSH and testosterone. In the female pubertal assay, the most variable endpoints included percent of females regularly cycling, ovarian, thyroid and pituitary weights, creatinine, and serum TSH concentrations. For some endpoints, the variability across laboratories was quite similar, which may suggest inherent variability in these endpoints. For other endpoints, interlaboratory differences in variability were noted. One contributor to the variability in organ weight measurements is prosector variability in collecting and trimming organs before weighing. Therefore, limiting the number of prosectors on a study may reduce organ weight variability, thereby increasing sensitivity for detecting compound-related effects for the assay endpoints.

Assay Interpretation: Systemic Toxicity and Body Weight Effects

Many endpoints included in the pubertal assays can be altered by changes in rate of growth and/or terminal body weight, making it difficult to interpret assay data and discern specific endocrine-mediated effects. There are some conflicting reports on the sensitivity of puberty onset to moderate changes in body weight/growth rate

Table 3
Interlaboratory Control Data for Male Pubertal Assay

Endpoint	Lab #1			Lab #2			Overall values		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
Age at PPS (days)	44.9	0.79	1.8	44.0	0.80	1.8	44.6	0.90	2.0
Body wt at PPS (g)	220.8	11.80	5.3	207.2	8.24	4.0	215.6	12.35	5.7
Age at incomplete PPS (days)	43.1	0.75	1.7	NA	NA	NA	43.1	0.75	1.7
BW initial (g)	56.9	3.02	5.3	55.2	2.04	3.7	56.2	2.77	4.9
Final BW (g)	285.9	8.28	2.9	283.4	8.21	2.9	285.0	8.14	2.9
Liver wt (g)	13.4	0.53	4.0	12.7	0.41	3.2	13.1	0.57	4.4
Kidney wt (g)	2.0	0.09	4.5	2.0	0.06	3.0	2.0	0.08	4.0
Pituitary wt (mg)	11.0	0.81	7.4	9.8	0.29	3.0	10.5	0.90	8.6
Adrenal wt (mg)	43.1	3.19	7.4	44.2	2.17	4.9	43.6	2.84	6.5
Ventral prostate wt (mg)	242.2	20.91	8.6	219.1	13.91	6.3	233.4	21.50	9.2
Dorsolateral prostate wt (mg)	117.4	9.46	8.1	166.6	13.60	8.2	136.1	26.81	19.7
Wt SV w/CG with fluid (mg)	645.5	54.96	8.5	495.6	61.12	12.3	588.4	93.17	15.8
Wt SV w/CG without fluid (mg)	406.0	20.20	5.0	259.4	48.82	18.8	347.4	80.88	23.3
LABC wt (mg)	528.4	26.55	5.0	463.7	77.25	16.7	503.8	59.57	11.8
Right epididymis wt (mg)	201.0	10.92	5.4	219.5	10.86	4.9	208.0	14.06	6.8
Left epididymis wt (mg)	196.4	10.06	5.1	214.2	8.68	4.1	203.2	12.85	6.3
Right testis wt (mg)	1444.1	75.78	5.2	1428.8	12.90	0.9	1438.3	59.69	4.2
Left testis wt (mg)	1420.5	65.10	4.6	1419.4	14.89	1.0	1420.1	51.20	3.6
Thyroid wt (mg)	13.6	1.90	14.0	11.8	0.64	5.4	12.9	1.77	13.7
Blood urea nitrogen (BUN) (mg/dl)	14.4	1.49	10.3	14.6	1.19	8.2	14.5	1.36	9.4
Creatinine (mg/dl)	0.1	0.07	70.0	0.2	0.05	25.0	0.15	0.06	40.0
Serum T4 levels (μ g/dl)	5.7	0.48	8.4	5.1	0.51	10.0	5.5	0.54	9.8
Serum TSH levels (ng/ml)	10.8	2.56	23.7	8.1	1.34	16.5	9.7	2.53	26.1
Serum T levels (ng/ml)	2.5	0.66	26.4	3.1	0.62	20.0	2.8	0.68	24.3

Lab #1: $n = 13$; Lab #2: $n = 8$.

PPS, preputial separation; BW, body weight; wt, weight; SV w/CG, seminal vesicles with coagulating glands; LABC, levator ani-bulbocavernosus muscles; T, testosterone.

(Table 5). Laws et al. (2007) reported that 20–21% decreases in body weight did not significantly affect age at puberty onset in male or female rats, suggesting age at puberty onset is insensitive to changes in growth; however, other studies suggest that age at puberty onset and body weight function as a continuum (Ashby and Lefevre, 2000), and body weight alterations of approximately 10–15% could alter puberty onset in male rats (e.g., Stoker et al., 2000; Marty et al., 2003; Carney et al., 2004). The differences reported in these publications may be related to the rate at which the body weight decrement occurred (i.e., how quickly it occurred and over what time frame/ages). Regardless of the cause for the differences in these reports, interpretation of the pubertal onset data can be confounded by numerous factors including body weight, and therefore caution must be taken in interpreting statistically significant effects to identify true endocrine-mediated effects from secondary effects due to systemic toxicity.

Organ weight measurements also may be affected by body weight decrements (body weight loss or slower rates of growth when compared to a control group). Across feed restriction studies, female organ weights were not altered with a 5% change in terminal body weight; however, the next level of feed restriction, which produced a 9% difference in terminal body weight, altered pituitary and kidney weights. The next level of feed restriction (12% difference in terminal body weight) altered adrenal, liver, and ovarian weights (Laws et al., 2007). Thus, a body weight change between 9 and 12% resulted in significant

differences in endocrine-related organ weights in female peripubertal rats. A 9% change in terminal body weight also altered the number of 4–5 day estrous cycles (Laws et al., 2000).

Feed restriction data for select endpoints in the male pubertal assay also are shown in Table 5. Overall, a 9–11% decrease in terminal body weight appears to be the threshold for significant decreases in several endocrine-sensitive organ weights in male pubertal rats, including epididymidal, ventral prostate, and seminal vesicle weights. Some organ weights, including adrenal, liver, kidneys, and pituitary, were altered with lower levels of feed restriction in male rats ($\leq 9\%$ difference in terminal body weight; Laws et al., 2007). There are no data as to whether the weights of the thyroid gland, LABC, or seminal vesicles without fluid are influenced by body weight changes as these organ weights have not been measured in pubertal assay feed restriction studies.

Data also indicate that feed-restricted males had a significant decrease in serum thyroxine (T4) and TSH levels with $\geq 9\%$ and 19% decreases in terminal body weight, respectively (Table 5) (Laws et al., 2007). Despite 10–25% decreases in T4 and TSH levels, none of the thyroid-related hormones were significantly altered in feed-restricted females at levels that resulted in a 19% decrease in terminal body weight. Serum testosterone levels were not significantly altered in feed-restricted male rats (Table 5).

Across studies, the precise magnitude of body weight change that results in significant differences in pubertal

Table 4
Interlaboratory Control Data for Female Pubertal Assay

Endpoint	Lab #1			Lab #2			Overall values		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
BW initial (g)	51.3	3.20	6.2	49.6	1.29	2.6	50.8	2.84	5.6
Final BW (g)	158.2	8.06	5.1	154.8	3.72	2.4	157.1	7.11	4.5
Age at VO (days)	35.4	0.99	2.8	34.6	1.02	2.9	35.2	1.05	3.0
Age at incomplete VO (days)	34.7	0.91	2.6	NA	NA	NA	34.7	0.91	2.6
Body wt at VO (g)	123.6	6.63	5.4	113.4	5.05	4.5	120.5	7.76	6.4
Age at first estrus (days)	36.8	0.95	2.6	35.6	1.00	2.8	36.4	1.11	3.0
Estrous cycle length (days)	4.7	0.27	5.7	4.7	0.18	3.8	4.7	0.24	5.1
Percent cycling	98.7	2.76	2.8	94.4	7.11	7.5	97.4	4.80	4.9
Percent regularly cycling	81.3	13.69	16.8	84.4	9.29	11.0	82.2	12.39	15.1
Liver wt (g)	7.4	0.63	8.5	7.2	0.21	2.9	7.3	0.54	7.4
Kidney wt (g)	1.3	0.08	6.2	1.3	0.04	3.1	1.3	0.07	5.4
Pituitary wt (mg)	9.7	1.03	10.6	7.9	0.20	2.5	9.1	1.20	13.2
Adrenal wt (mg)	34.8	1.78	5.1	35.5	3.85	10.8	35.0	2.52	7.2
Ovarian wt (mg)	74.6	4.87	6.5	57.5	3.80	6.6	69.4	9.21	13.3
Uterine wet wt (mg)	302.6	33.04	10.9	315.5	12.70	4.0	306.5	28.72	9.4
Uterine blotted wt (mg)	256.9	22.19	8.6	280.4	17.66	6.3	264.0	23.30	8.8
Thyroid wt (mg)	9.3	1.32	14.2	8.8	0.33	3.8	9.2	1.13	12.3
Blood urea nitrogen (BUN) (mg/dl)	13.2	1.25	9.5	12.7	1.11	8.7	13.1	1.21	9.2
Creatinine (mg/dl)	0.1	0.05	50.0	0.1	0.02	20.0	0.12	0.04	33.3
Serum T4 levels ($\mu\text{g}/\text{ml}$)	4.4	0.27	6.1	3.8	0.25	6.6	4.2	0.37	8.8
Serum TSH levels (ng/ml)	6.4	1.60	25.0	3.6	0.82	22.8	5.5	1.92	34.9

Lab #1: $n = 16$; Lab #2: $n = 7$.

BW, body weight; VO, vaginal opening; NA, not applicable; wt, weight.

assay endpoints is unclear; however, the female pubertal assay results must be interpreted with caution if a $\geq 10\%$ change in terminal body weight gain is observed, as stated in the U.S. EPA test guidelines (U.S. EPA, 2009a, 2009b). The male pubertal assay test guideline is more conservative on this point as the test guideline cautions that a 6% decrease in body weight gain at termination should be interpreted with caution using a WoE approach with other available information or that additional studies may be needed to determine endocrine activity. When considering the potential impact of terminal body weight on assay endpoints, it also is important to consider the difficulty in titrating dose level selection such that a significant decrease in body weight gain is achieved without exceeding a 10% difference from controls in terminal body weight, particularly in growing animals. In practice, decrements in terminal body weights that exceed 10% of the control group are likely to occur in some studies.

Assay Interpretation: Variability in Developmental Landmark Data

A primary endpoint of the pubertal assays is age and body weight at puberty onset (PPS and vaginal opening). Although each laboratory trains their technical staff according to detailed standard operating procedures, these developmental landmarks show some variability both among and within testing facilities. For example, in EPA's prevalidation and validation studies, mean age at PPS in control animals varied from 39.6 to 43.9 days of age and mean age at vaginal opening in control animals ranged from 31.5 to 34.9 days of age. The basis for this inter-animal variability in age at puberty onset is poorly understood, although puberty onset can be influenced by many

factors such as growth hormone, melatonin, higher brain function, diet composition, etc. (e.g., Frisch et al., 1975; Zipf et al., 1978; Smith et al., 1989; Cicero et al., 1990, 1991; Odum et al., 2001). In addition, since attainment of puberty is a subjective evaluation, interlaboratory, and even intralaboratory variability will be a factor.

Variability across laboratories has been a concern with regard to the evaluation of pubertal developmental data for various compounds. It is difficult to ascertain clear agreement amongst laboratories when evaluating the criteria for achievement of vaginal patency and PPS; however, it is recommended that the same personnel evaluate all animals on a given study whenever possible to minimize intralaboratory variability. Comparison of HCD maintained by several laboratories reinforces this concern regarding cross-laboratory variability (Table 6 and Figs. 2 and 3). Developmental landmark HCD in CD rats (Sprague-Dawley) were compiled from three separate Good Laboratory Practices (GLP)-compliant laboratories within the last 5 years. The data were divided into reproductive toxicity studies and EDSP studies (i.e., pubertal male and female assay data) to determine if there was any inter- and intralaboratory variability observed based on study type. For reproductive toxicity studies, the historical control mean age of achievement for vaginal patency ranged from 31.9 to 33.2 days and the age range for PPS was 43.4–46.3 days across laboratories. The historical control mean age of achievement for EDSP studies ranged from 32.8 to 35.6 days for vaginal patency and 42.2–45.0 days for PPS. The reason for this disparity between reproductive toxicity studies and EDSP studies is not clearly understood, although the larger sample sizes evaluated in reproductive toxicity studies may give these data greater precision.

Table 5
Effects of Feed Restriction on Select Endpoints in the Pubertal Female and Male Assays:
Terminal Body Weight, Age at Puberty Onset, and Organ Weights

Female pubertal assay parameters	Laws et al. (2007) Change from control	Laws et al. (2007) Change from control	Laws et al. (2000) Change from control	Laws et al. (2007) Change from control	Laws et al. (2007) Change from control
Strain	Wistar	Wistar	Wistar	Wistar	Wistar
Sample size (<i>n</i>)	13	13	7–8	13	13
Age at start	22	22	22	22	22
Age at termination	41–42	41–42	41	41–42	41–42
Terminal bwt decrease	–2.1%	–4.6%	–8.6%	–12.1%	–18.9%
Age at vaginal opening (VO) (difference in days)	+0.3	–0.4	+1.6	+1.0	+1.5
Body weight at VO	+1.7%	–2.8%	–4.9%	–2.9%	–9.1%
No. of 4–5 day cycles (monitored: VO + 15 days)	NA	NA	–40.7% ^a	NA	NA
Pituitary	–2.8%	–2.6%	–16.3%	–13.3%	–21.6%
Adrenal	–2.5%	–7.1%	–1.0%	–13.6%	–17.0%
Liver	–0.7%	–13.4%	–5.1%	–16.8%	–29.5%
Kidneys	–2.7%	–7.4%	–7.5%	–13.4%	–20.8%
Ovaries	–6.7%	–3.3%	–11.4%	–21.7%	–31.7%
Uterus with fluid	–10.3%	–18.7%	–25.1%	–13.7%	–32.1%
Uterus without fluid	–3.2%	–6.0%	–12.6%	–10.8%	–28.3%
T4	+13.5%	+4.7%	–0.5%	–8.6%	–9.7%
TSH	+9.0%	–16.3%	+35.3%	–17.5%	–24.5%

Male pubertal assay parameters	Laws et al. (2007) Change from control	Laws et al. (2007) Change from control	Laws et al. (2007) Change from control	Marty et al. (2003) ^b Change from control	Stoker et al. (2000) Change from control	Laws et al. (2007) Change from control
Strain	Wistar	Wistar	Wistar	CD	Wistar	Wistar
Sample size (<i>n</i>)	13	13	13	12	10 ^c	13
Age at start	23	23	23	23	23	23
Age at termination	53–54	53–54	53–54	52	53	53–54
Terminal bwt decrease	1.8% ^d	5.9% ^d	9.0% ^d	11.3%	15.0%	19.2% ^d
Age at preputial separation (PPS) (difference in days)	–0.5	+0.3	–1.4	+1.8 ^e	+2.1 ^f	–0.1
Body weight at PPS	+1.2%	–1.5%	–11.6%	–5.4%	–13.4%	–16.5%
Pituitary	–5.1%	–12.8%	–17.4%	NA	NA	–23.4%
Adrenals	–17.6%	11.6%	–18.5%	NA	NA	–26.8%
Liver	–8.8%	–14.5%	–22.0%	NA	NA	–32.8%
Kidneys	–6.9%	–11.5%	–15.6%	NA	NA	–27.1%
Seminal vesicles with fluid	+17.3%	–2.7%	–11.4%	–10.5%	–55.2%	–30.7%
Ventral prostate	+10.7%	–6.7%	–12.6%	–11.4%	–29.2%	–23.3%
Epididymis (left)	+0.4%	–3.6%	–3.2%	–8.8% ^g	ND ^h	–10.4 ^g
Epididymis (right)	NA	NA	NA			NA
Testis (left)	+2.1%	–1.4%	–2.1%	–4.6% ⁱ	ND ^j	–4.9%
Serum testosterone	+42.9%	–31.1%	+43.5%	NA	+9.5%	–5.0%
Serum T4	–3.2%	–14.4%	–23.2%	NA	+16.6%	–25.8%
Serum TSH	–10.9%	+1.4%	–15.2%	NA	–17%	–31.2%

NA, not applicable; endpoint not measured.

Bold type indicates changes in body weights, age at preputial separation, organ weights, or hormone concentrations that were statistically significant due to feed restriction and decreased body weights.

^aNo. of 4–5 day cycles from VO to 15 days was 2.7 in ad libitum fed control animals versus 1.6 in feed-restricted animals.

^bValues derived from animals necropsied on PND 52, the time point closest to the current study necropsy on PND 53.

^c*n* = 10 in the feed-restricted group, whereas *n* = 24 in the ad libitum fed control group.

^dBody weight differential on PND 53, although rats were euthanized on PND 53–54.

^eBody weight differential on PND 41 was approximately 13% relative to ad libitum fed controls.

^fBody weight differential on PND 43 was approximately 20% relative to ad libitum fed controls.

^gPercent change is for paired epididymal weights.

^hNo data available: significant difference in absolute epididymides weight reported, but data not shown.

ⁱPercent change is for paired testis weights.

^jNo data available: no significant difference in absolute testes weights reported, but data not shown.

Table 6
Historical Control Data for Puberty Onset Endpoints in Sprague-Dawley Rats for Studies Conducted from 2007 to 2013^{a,b}

	Company A EDSP studies	Company A Reproductive studies	Company B EDSP studies	Company B Reproductive studies	Company C EDSP studies	Company C Reproductive studies	Performance criteria (OPPTS 890.1500 and 890.1450)
Preputial separation							
Age (days)	45.0 (0.9)	45.0 (1.3)	42.2 (0.9)	46.3 (1.2)	44.0 (0.80)	44.6 (0.58)	43.1 (39.8–46.5) ^c
Body weight at achieve- ment	220.7 (12.1)	231.2 (12.4)	224.6 (10.8)	239.0 (12.2)	207.2 (8.2)	245.9 (10.1)	222.2 (188.277–256.169) ^d
Vaginal patency							
Age (days)	35.6 (1.1)	33.2 (1.0)	32.8 (1.2)	32.8 (0.9)	34.6 (1.0)	32.4 (1.0)	33.2 (30.7–35.6) ^c
Body weight at achieve- ment	124.1 (7.7)	110.6 (5.5)	119.6 (9.6)	105.4 (7.0)	113.4 (5.0)	116.2 (5.9)	116.6 (101.71–131.44) ^d
Sample size (no. of studies)	9	165	3	86	7–8	8	

^aComparison of laboratory mean values versus performance criteria means (ranges) as specified in the pubertal assay test guidelines.

^bValues are means (standard deviations) for each laboratory and study type.

^cAge at preputial separation or vaginal opening in Sprague-Dawley rats based on the performance criteria specified in the OPPTS 890.1450 and 890.1500 test guidelines.

^dBody weight at preputial separation or vaginal opening in Sprague-Dawley rats based on the performance criteria specified in the OPPTS 890.1450 and 890.1500 test guidelines.

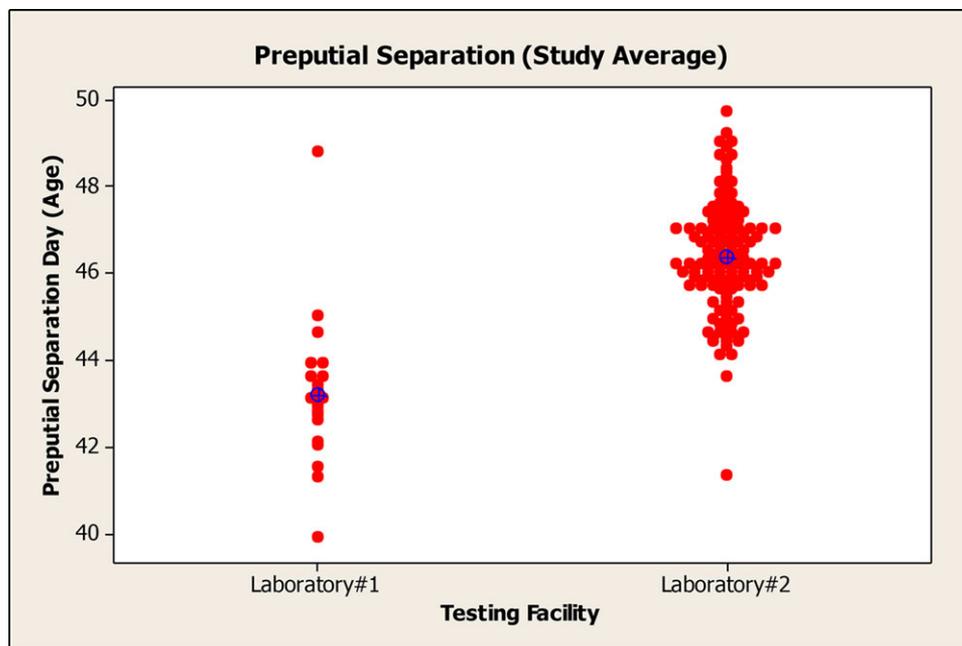


Fig. 2. Distribution of age at preputial separation across two laboratories, which shows slight interlaboratory variability in the measurement of puberty onset in males.

Furthermore, there are other inherent differences in the study design that may contribute to this variability, including type of diet (Odum et al., 2001; You et al., 2002), cage bedding (Markaverich et al., 2002), and daily gavage administration in the pubertal studies, which may be more stressful than dietary dosing in reproductive toxicity studies. Both the reproductive toxicity study data and EDSP pubertal data were within the acceptable age and body weight ranges for puberty onset as specified by the performance criteria for

the male and female pubertal assays (OPPTS 890.1450 and 890.1500). However, it may be inappropriate to use HCD from pubertal studies to aid the interpretation of age of puberty onset data from reproductive toxicity studies.

Assay Interpretation: Estrous Cycle Data

Determining a correlation between altered endocrine-sensitive endpoints and a primary endocrine MoA from

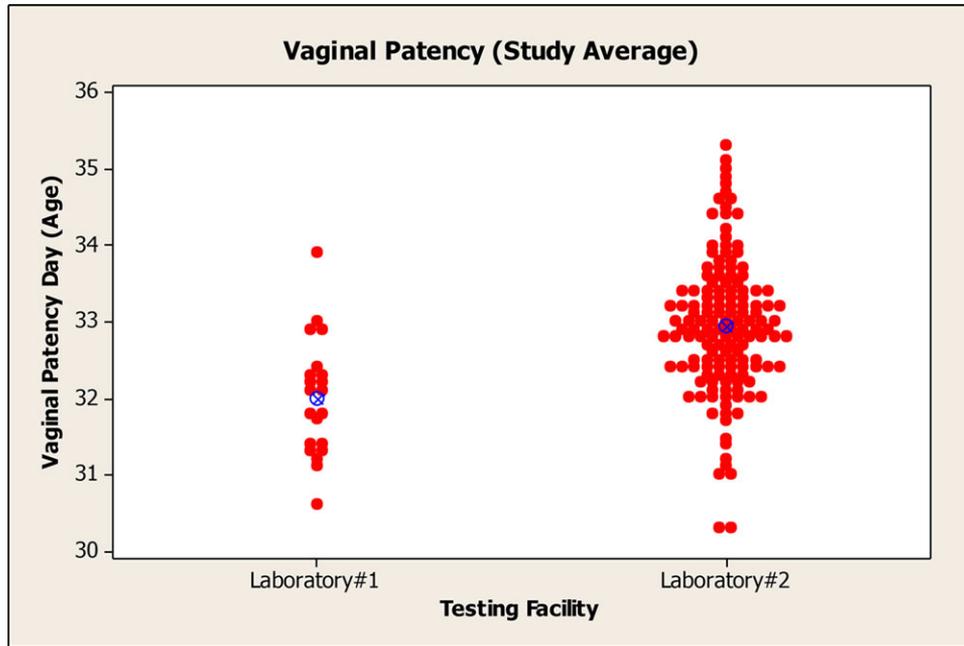


Fig. 3. Distribution of age at vaginal opening across two laboratories. There was minimal interlaboratory variability in the distribution of age at puberty onset in females.

the test compound can be challenging in the pubertal assays. For example, one of the major challenges for the female pubertal assay is the assessment of estrous cyclicity. The mean day of attainment of vaginal patency typically occurs approximately 10 days before the necropsy of the females on PND 42 and a normal estrous cycle length in rats is 4–5 days. If monitoring begins mid-cycle, it may take 8 days or longer to observe two estrus stages to determine estrous cycle length. The test guideline requires each female to be characterized as “regularly cycling,” “irregularly cycling,” or “not cycling”; however, the monitoring interval may not allow for the evaluation of a full estrous cycle, particularly if an animal is slightly older at the time of vaginal opening (acceptable mean range in control animals is 30.67–35.62 days of age). In these cases, it may not be possible to determine if an animal is cycling normally, because the monitoring period is too short. In addition, there are interanimal differences in the duration of estrous cycle monitoring such that monitoring across the dose groups is often inequitable. To perform a thorough assessment of estrous cyclicity requires a 2–3 week period as required in the multigeneration reproduction study design.

The interpretation of estrous cycle data is further complicated by inherent variability in cycle length or pattern, particularly with the onset of cycling. It is not uncommon for young animals to cycle abnormally with the initiation of estrous cycling (it usually takes until about 8 weeks of age for normal cycles to occur consistently) and the estrous cycle also can be influenced by other factors such as stress and feed intake (Matysek, 1989; Roozendaal et al., 1995; Laws et al., 2000). In prevalidation work for the female pubertal assay, one study had 12 of 14 control animals that failed to achieve regular cycles during the monitoring period after vaginal opening (U.S. EPA, 2007a). Fur-

thermore, an 8.6% difference in terminal body weight has been shown to decrease the number of 4–5 day cycles after vaginal opening (Laws et al., 2000). The U.S. EPA stated that the “EPA recognizes that estrous cyclicity may not be well established within the duration of the pubertal assay even in control animals and thus will generally not rely on small deviations as contributing heavily to the weight of evidence” (U.S. EPA, 2011b). Thus, estrous cycle data should be evaluated with caution and used primarily to support other evidence of altered endocrine function.

Furthermore, the test guideline uses a conservative description of normal estrous cycle patterns, stating that “estrous cycle length” is from the first day of one proestrus to the first day of the next proestrus (or first day of diestrus or estrus to the next first day of diestrus or estrus; U.S. EPA, 2011b) and that an animal is “irregularly cycling” if it has a period of diestrus longer than 3 days (or a period of cornification longer than 2 days). In rats, the proestrus stage of the estrous cycle is 12–24 hr in duration (Zarrow et al., 1964); thus, with once daily vaginal smears, it is possible to miss this stage of the cycle. Consequently, animals exhibiting a cycle with the pattern “estrus-diestrus-diestrus-diestrus-estrus” are equivalent to animals having the pattern “estrus-diestrus-diestrus-diestrus-proestrus-estrus” if proestrus was missed; however, according to the test guideline, the first pattern is irregular, whereas the second pattern is normal. The regularity of “estrus-diestrus-diestrus-diestrus-estrus” cycling has been confirmed with HCD from one of the participating laboratories, where estrous cycles were monitored for 4 weeks in young control CD rats (PND 40–68). In this data set, 10 of 27 animals had several intervals with four consecutive days of diestrus, but exhibited regular 4- or 5-day cycles over the monitoring period. Thus, if these slightly older animals (PND

Table 7
Thyroid Histopathology Using a 5-Point Grading Scale

<u>Follicular cell height</u>	
Grade 1	Prominent number of follicles (>25%) were lined by an attenuated epithelium with scant eosinophilic cytoplasm and a flattened hyperchromatic nucleus. The majority of the remaining follicles were lined by a low cuboidal epithelium with a round to oval nucleus and a cell height that was $\leq 1.5\times$ the height of the nucleus.
Grade 2	Most follicles were lined by a cuboidal to slightly columnar epithelium, often with a flocculent to finely vacuolated eosinophilic cytoplasm, round to oval nucleus, and the cell height which was $\leq 2\times$ (twice) the height of the nucleus.
Grade 3	Most follicles were lined by a slightly columnar epithelium with a foamy to vacuolated eosinophilic cytoplasm and a round nucleus; the cell height is $2\times$ to $2.5\times$ the height of the nucleus.
Grade 4	Similar nuclear and cytoplasmic characteristics as described for Grade 3 above. Most follicles were lined by a distinctly columnar epithelium with a cell height $2.5\times$ to $3\times$ the height of the nucleus.
Grade 5	Most follicles were lined by a distinctly columnar epithelium with a cell height $>3\times$ the height of the nucleus.
<u>Colloid area</u>	
Grade 1	Absence of colloid or decreased colloid in $>67\%$ of thyroid follicles.
Grade 2	Same as Grade 1 above, except the condition affected 34–66% of the follicles.
Grade 3	Most follicles were small, even at the periphery, with 25–33% being collapsed with no visible colloid or decreased amount of colloid.
Grade 4	Most follicles contained variable amounts of eosinophilic to grayish-pink colloid with the peripheral follicles generally being larger with more abundant colloid compared to the innermost follicles.
Grade 5	Most follicles were filled with eosinophilic colloid with only a slight variation in follicular size.

40–68), with more stable cycles, exhibit these estrous cycle patterns, it would be inappropriate to label juvenile animals, with less stable cycles, as “irregular” or “non cycling” when they show the same estrous patterns.

Assay Interpretation: Female Organ Weight Data with Terminal Estrous Stage

Aside from estrous cycle data, estrous stage at necropsy is important to consider when evaluating organ weights in pubertal females, particularly for uterine and ovarian weights. Furthermore, the variability in terminal stage of estrous imparts greater variability on these organ weights; for example, in pubertal female studies conducted in one author’s laboratory, uterine weight for females in proestrus on the day of necropsy were up to threefold higher than uterine weights for females in diestrus from the same dose group on the day of necropsy. Therefore, it is important to consider stage of estrous at termination when interpreting organ weight differences across treatment groups.

Assay Interpretation: Thyroid Endpoints

The male and female pubertal assays introduced the requirement to collect thyroid histopathology data (follicular cell height and colloid amount) using a new “1 to 5” grading scale to aid in the identification of thyroid histopathological changes. The test guidelines include photomicrographs designed to illustrate the various grades of thyroid histopathological changes. When applying this system in the laboratory, the thyroid grading scale described in Table 7 was used in one of the authors’ laboratories.

At this time, it is unclear whether the collection of thyroid histopathology data using a “1 to 5” scale aids in the identification of thyroid histopathological changes. Thyroid scaling may have been introduced to improve objectivity in histopathological assessments; however, histopathological evaluations are generally qualitative assessments and may not lend themselves to the precision

implied by numeric scaling. Any changes in thyroid follicular cell height or colloid amount should be interpreted in conjunction with thyroid hormone levels because effects on these parameters are typically reversible if the effects on serum T4 and TSH are not sustained. In at least one case from the authors’ laboratories, a compound was designated as altering thyroid histopathology despite the lack of a statistically significant difference in scaled thyroid values. In this case, the professional judgment of experienced pathologists had more utility than quantitative thyroid values.

Assay Interpretation: Hormone Data

The male and female pubertal assays require the determination of serum T4 and TSH concentrations, as well as serum testosterone concentrations in males. Caution is warranted when interpreting changes in thyroid hormones (TSH, T4) without corresponding changes in thyroid weights or histopathology. Thyroid hormone levels represent a measurement at a single point in time (necropsy), whereas thyroid weight and histopathology are endpoints that represent cumulative events. Differences in thyroid hormone concentrations may be related to other factors such as stress at the time of necropsy (Döhler et al., 1979), estrous stage (Döhler et al., 1979), decreased body weight/body weight gain (Laws et al., 2007), or fasting/nutritional status of the animals (Eales, 1988; Boelen et al., 2008). The U.S. EPA has advocated that “the biological/toxicological significance of changes in thyroid hormone levels in the absence of corroborative histopathological changes will be evaluated in the context of the overall toxicity of the compound using the WoE approach including the thyroid toxicity data available from the amphibian metamorphosis assay” (U.S. EPA, 2011b). In these cases, consistent thyroid changes between the male and female pubertal assays and/or supporting evidence from the amphibian metamorphosis assay may aid in the determination of specific thyroid effects.

Serum testosterone concentrations also are subject to change due to indirect effects on the endocrine system. With respect to serum testosterone levels, the FIFRA Scientific Advisory Panel (FIFRA SAP, 2013) noted that low level, chronic stress can significantly decrease testosterone secretion as well as contribute to increased variability in serum hormone values.

PRACTICAL EXPERIENCE ON ENDPOINT SPECIFICITY AND SENSITIVITY

In the male pubertal assay, there are numerous endpoints that are sensitive to changes in androgen signaling, including puberty onset, reproductive and accessory sex tissue weights, reproductive organ histopathology, and serum testosterone concentrations. In a similar fashion, the female pubertal assay has several endpoints that can be altered in response to modulation of estrogen signaling (e.g., puberty onset, age at first estrus, estrous cyclicity, reproductive organ weights, and histopathology). Alterations in the thyroid pathway can be detected in both the male and female pubertal assays by monitoring thyroid weight, histopathology, and serum levels of TSH and T4. Ideally, patterns of effects across multiple endpoints could be used to discern which pathway is affected and possibly, the MoA by which changes occur. This section examines data generated in numerous pubertal assays to examine, in practical terms, the relative sensitivity of related endpoints and the potential to generate patterns of effects expected for an endocrine-active compound.

Using data from 21 male pubertal assays and 23 female pubertal assays, Tables 8 and 9 summarize the frequency of statistically significant endpoints. In these tables, a "1" is used to identify a statistically significant finding, whereas a "0" indicates that the endpoint was not statistically different from control values. Although not statistically analyzed, histopathology endpoints were added to the bottom of Tables 8 and 9 to show where treatment-related changes were identified by the study pathologist. It is important to note that thyroid follicular cell height and colloid area were scored (i.e., 1–5) and statistically analyzed, whereas other histopathology evaluations (thyroid, testis, epididymis, kidney, ovaries, and uterus) were qualitative evaluations.

Aside from the histopathology endpoints, all other endpoints were altered in one or more studies with the exception of relative pituitary weight in the male pubertal assays, which was not affected in any of the current studies. Most male and female assay endpoints were altered in 4–38% of studies except for relative liver weights, which were altered in 52–57% of studies in both genders (Tables 8 and 9). In seven of 12 studies (male) and 6 of 12 studies (female), the liver weight changes occurred in the absence of significant body weight/body weight gain changes. It is plausible that relative liver weights were often increased due to enzyme induction secondary to daily bolus dosing of test materials. The Society of Toxicology's Task Force to Improve the Scientific Basis of Risk Assessment (Conolly et al., 1999) identified gavage as an "unrealistic method" of exposure due to the potentially rapid delivery rate of the test material to the target site; however, the conservative nature of the EDSP, which was de-

signed to minimize false-negative findings in Tier 1, may justify use of this route of exposure to identify compounds for further evaluation. Notably, Tier 1 assays were originally not intended for risk assessment purposes, so the use of gavage exposures provided a conservative screening tool.

As seen in Tables 1 and 2, the endpoint most commonly used to establish an MTD in the male and female pubertal assays was body weight/body weight gain, which was targeted in 8 of 20 compounds in the male assays and 10 of 23 compounds in the female assay. When the assays were actually conducted, body weight or body weight gain was significantly altered in approximately 32% of male and female pubertal studies (six male assays and eight female assays; Tables 8 and 9). These data suggest that range-finding studies are useful for selecting MTDs; however, the range-finding studies are not 100% predictive of pubertal assay outcomes. It is unclear whether these deficiencies in the predictiveness of the range-finding studies were related to the study designs used (e.g., smaller sample sizes, inadequate monitoring period, etc.) or variability in the responsiveness of juvenile animals.

Kidney weights were significantly altered in 29% of the male assays and 26% of the female assays. In five of six studies (both male and female assays), kidney weight changes occurred in the presence of liver weight changes, but did not necessarily correspond with body weight changes.

With regard to male androgenic endpoints, LABC weight was the most frequently affected endpoint. This muscle weight was affected in eight (38%) of the studies. Ventral prostate and seminal vesicle weights also were typically altered in conjunction with LABC weights. Androgenic endpoints affected in 24–33% of studies included absolute and adjusted weights of the ventral prostate, seminal vesicles with coagulating glands (with and without fluid), right and left epididymides, and dorsolateral prostate (adjusted weight based on PND 21–23 body weight). Testicular weights were rarely affected. Serum testosterone levels were significantly altered in 29% of the male pubertal studies. Typically, seminal vesicle weights with and without fluid showed the same result when statistically analyzed. Also, the results for adjusted and absolute values were generally in agreement, although there were 13 cases in the male and female pubertal assays in which either the absolute or adjusted value was significant, but not both values. This was especially problematic for age at PPS. Puberty onset was altered in less than 20% of studies.

Patterns of effects were present when examining across androgen-sensitive endpoints. In eight studies, three-to-six androgen-sensitive endpoints were significantly altered. For example, effects on the LABC muscles generally occurred in the presence of alterations in other accessory sex tissue weights, including effects on seminal vesicle (with and without fluid), prostate (ventral and dorsolateral), and/or epididymal weights. However, in seven of these eight cases, clinical chemistry parameters (creatinine and/or BUN) and/or body weight/body weight gain also were altered, indicating that an MTD may have been exceeded. In these cases, the interpretation of these findings is confounded by these other variables, such that the specificity of an

Table 8
Male Pubertal Assays: Frequency of Endpoint Alterations

Endpoint	Study data: statistically altered endpoint (0 = not altered; 1 = altered)																				Statistically altered endpoint (%)
	A ^a	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
Body wt ^b /gain (prepuberty) ^c	0	1	0	0	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	29
Body wt/gain (termination)	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	24
Absolute age at preputial separation (PPS)	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	19
Adjusted age at PPS	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	19
Absolute body wt at PPS	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	10
Adjusted body weight at PPS	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	10
Relative liver wt	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	57
Relative kidney wt	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	1	0	29
Relative pituitary wt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Relative adrenal wt	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	10
Absolute ventral prostate wt	0	0	0	1	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	29
Adjusted ventral prostate wt	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	0	0	0	0	0	33
Absolute dorsolateral prostate wt	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	24
Adjusted dorsolateral prostate wt	0	1	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	29
Absolute wt SV w/CG with fluid ^d	1	0	0	1	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	33
Adjusted wt SV w/CG with fluid	1	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	33
Absolute wt SV w/CG without fluid	1	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	29
Adjusted wt SV w/CG without fluid	1	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	33
Absolute LABC wt ^e	0	1	0	1	1	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	38
Adjusted LABC wt	0	1	0	1	1	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	38
Absolute right epididymis wt	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	29
Adjusted right epididymis wt	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	29
Absolute left epididymis wt	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	24
Adjusted left epididymis wt	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	24
Absolute right testis wt	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	10
Adjusted right testis wt	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	5
Absolute left testis wt	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Adjusted left testis wt	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Absolute thyroid wt	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	10
Adjusted thyroid wt	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	10
Blood urea nitrogen (BUN)	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	19
Creatinine	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	24
Serum T4 levels	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	19
Serum TSH levels	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	14
Serum T levels	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	29
Thyroid follicular cell height	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	10
Thyroid colloid amount	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	10
Thyroid histopathology ^f	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	24
Testis histopathology	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Epididymis histopathology	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Kidney histopathology	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	14

^aData in this table reflect the same studies referenced in Table 1, although the sequence has been changed.

^bwt, Weight.

^cBody weight/body weight gain during intervals leading up to PPS (e.g., PND 35–45).

1 = Statistically significant difference in treated group(s) compared with concurrent controls; 0 = no difference between treated and control groups.

^dSV w/CG, seminal vesicles with coagulating glands.

^eLABC, levator ani-bulbocavernosus muscles.

^fHistopathology results for thyroid, testis, epididymis, and kidney were qualitative; therefore, 1 = treatment-related difference and 0 = no difference from controls.

endocrine-mediated effect may be difficult to determine. Previous feed restriction studies have verified that reproductive and accessory sex tissue weights can be affected by alterations in body weights/body weight gains in the male pubertal assay (see discussion above). Interestingly, there also was one study in which changes in LABC occurred in the absence of changes to any other androgen-sensitive endpoints. The significance of this finding is unknown.

Serum testosterone levels, which were altered in 29% of studies, were sometimes supportive of changes in androgen-sensitive tissues. In three of the eight studies in which multiple androgen-dependent organ weights were altered, serum testosterone was significantly affected. In a separate case, altered seminal vesicle weights and increased adjusted age at PPS were seen in the presence of changes in testosterone levels; however, there were no effects on other reproductive or accessory sex tissue

Table 9
Female Pubertal Assays: Frequency of Endpoint Alterations

Endpoint	Study data: statistically altered endpoint (0 = not altered; 1 = altered)																				Statistically altered endpoint (%)			
	A ^a	B	C	D	E	F	G	H	I ^e	J	K	L	M	N	O	P	Q	R	S	T		U	V	W
Body wt ^b /gain (prepuberty) ^c	0	1	0	0	1	0	1	0	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	30
Body wt/gain (termination)	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	13
Absolute age at vaginal opening (VO)	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	17
Adjusted age at VO	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	17
Absolute body wt at VO	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9
Adjusted body weight at VO	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9
Age at first estrus	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Estrous cycle length	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Percent cycling	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Percent regularly cycling	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Relative liver wt	1	1	0	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	1	52
Relative kidney wt	0	0	0	0	0	0	1	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	1	26
Relative pituitary wt	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	17
Relative adrenal wt	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4
Absolute ovarian wt	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	13
Adjusted ovarian wt	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	17
Absolute uterine wet wt	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	13
Adjusted uterine wet wt	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	13
Absolute uterine blotted wt	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9
Adjusted uterine blotted wt	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	17
Absolute thyroid wt	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Adjusted thyroid wt	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Blood urea nitrogen (BUN)	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	13
Creatinine	1	0	0	1	1 ^f	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	17
Serum T4 levels	0	1	0	0	1	0	1	NA	NA	0	0	1	0	0	0	0	1	0	0	0	0	0	0	24
Serum TSH levels	0	1	0	0	1	0	1	NA	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
Thyroid follicular cell height	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13
Thyroid colloid amount	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
Thyroid histopathology ^d	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13
Ovarian histopathology	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Uterine histopathology	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Kidney histopathology	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	9

^aData in this table reflect the same studies referenced in Table 1, although the sequence has been changed.

^bwt, Weight.

^cBody weight/body weight gain during intervals leading up to vaginal opening (e.g., PND 28–33).

1 = Statistically significant difference in treated group(s) compared with concurrent controls; 0 = no difference between treated and control groups.

^dHistopathology results for thyroid, ovary, uterus, and kidneys were qualitative; therefore, 1 = treatment-related difference and 0 = no difference from controls.

^eAntiestrogenic

^fNo corresponding kidney histopathology or consistent changes with BUN. NA = not applicable

weights and no effect on absolute age at PPS. There were five studies with significant changes in reproductive and accessory sex tissues that did not identify significant differences in serum testosterone levels. Furthermore, there also was one study in which serum testosterone was significantly affected without any changes in any other androgen-dependent endpoints. Thus, serum testosterone levels may be more useful as supportive evidence of other effects in androgen-sensitive endpoints, but should not be the primary determining factor for or against potential androgenic/antiandrogenic activity.

With respect to puberty onset, absolute and adjusted age at PPS were each significantly altered in four studies; however, in only two studies were both absolute and adjusted age at PPS altered together. In the two studies where absolute age at PPS was altered and adjusted age was not, effects on body weight/body weight gain

were seen, which could confound the interpretation of this endpoint. Changes in the adjusted age at PPS (adjusted for weanling body weight before dosing) were seen in the absence of significant body weight changes. In any event, changes in age at PPS were affected less frequently than reproductive organ weights and never in the absence of effects on other androgen-sensitive endpoints in the current data sets. Body weight at PPS was significantly affected in two studies; however, in one of these studies, there were no other alterations in endocrine-sensitive endpoints. Overall, in these data sets, age at PPS was not the most sensitive endpoint to detect potential androgen-related effects.

Testicular histopathology was not affected in any of the pubertal male studies and epididymal histopathology was only affected in one study. Thus, it is unclear whether the compounds tested in these data sets pose a hazard

to subsequent reproductive tissue function or whether homeostasis might be reestablished without long-term adverse effects.

For thyroid endpoints, male serum TSH and T4 levels were altered in 14 and 19% of studies, respectively. It is not surprising that TSH was altered in a smaller proportion of studies, because TSH is a more variable endpoint than serum T4 levels. In two of three studies in which TSH was affected, serum T4 also was altered. In one study in which both TSH and T4 changes were detected, thyroid histopathology was significantly altered, whereas in the other study, both thyroid weight and thyroid histopathology were altered. Serum T4 was altered in two studies with no corresponding changes in serum TSH, thyroid weight, or histopathology (see discussion above on the specificity of changes in serum T4). Interestingly, thyroid histopathological changes were detected with one compound that did not significantly affect serum T4 levels. It is possible that the animals adapted to initial thyroid perturbations by this compound, and reestablished thyroid hormone homeostasis through follicular cell hypertrophy. Some compounds that altered thyroid hormones also produced effects on body weight/body weight gain, which may confound data interpretation (Laws et al., 2007).

There is a question related to the sensitivity of thyroid-related endpoints in the male pubertal assay as a weak-acting thyroid agent (e.g., phenobarbital) did not alter thyroid weights, thyroid histopathology, or serum T4 or TSH levels in the male pubertal assay (U.S. EPA, 2007b). Instead, phenobarbital delayed PPS, and decreased reproductive and AST weights, producing a pattern of effects similar to antiandrogens such as linuron and flutamide. Phenobarbital can affect serum LH levels (O'Connor et al., 1999b); however, it also is a known hepatic enzyme inducer; therefore, enhanced steroid hormone metabolism may have caused the antiandrogenic signals seen in the male pubertal assay. In one of the authors' laboratories, compounds detected for potential antiandrogenicity in the Hershberger assay due to enhanced metabolism of testosterone propionate, also produced antiandrogenic responses in the male pubertal assay, despite the absence of effects on androgen receptor (AR) binding or steroidogenesis by these compounds. Thus, it seems possible that enzyme-inducing compounds that enhance testosterone metabolism may produce responses in the male pubertal assay that appear to indicate antiandrogenicity, when the responses simply reflect hepatic enzyme induction.

With regards to female estrogenic endpoints, statistical significance was observed on the age of vaginal opening in four (17%) of the studies. In three of the studies with an effect on age at vaginal opening, there were other correlating effects on estrogenic endpoints such as age of first estrus, estrous cycle length, percentage of cycling/regularly cycling animals, ovarian and uterine weight, and/or ovarian and uterine histopathology. In one study, a delay in age at vaginal opening was observed without any other effect on estrogen-related parameters. The conclusion in this study was that the delay in vaginal opening was secondary to a statistically significant lower body weight gain. Body weight and/or body weight gain were significantly altered in three of the four studies in which puberty onset was affected, including one compound that was identified as an antiestrogen. One compound that

specifically altered puberty onset parameters and uterine weight (without any body weight changes) has been identified as an aromatase inhibitor.

Interestingly, an aromatase inhibitor that was evaluated by one of the authors did not alter any estrogen-sensitive endpoints in the female pubertal assay. The sensitivity of the female pubertal assay to detect weak aromatase inhibitors has been questioned (Marty et al., 1999; U.S. EPA, 2007a). In prevalidation, the weak-to-moderate aromatase inhibitors, fenarimol and δ -testolactone, were not detected in the female pubertal assay (Marty et al., 1999; U.S. EPA, 2007a), whereas the potent aromatase inhibitor, fadrozole, was readily detected in the female pubertal assay (Marty et al., 1999). The U.S. EPA may provide greater clarity on this issue as more compounds are evaluated. Most importantly, aromatase inhibitors can be detected by other assays in the Tier 1 battery.

Reproductive organ weights can be difficult to interpret in female animals, when stage of estrous is not controlled at necropsy. In three studies, ovarian weight changes were observed in the absence of any other changes in estrogenic endpoints; these ovarian weight changes occurred in the presence of changes in body weight/body weight gain. There were no associated ovarian histopathological findings. These data suggest that these ovarian weight changes were not the result of an endocrine-mediated change. In three studies, there was a good correlation between effects on wet and/or blotted uterine weights with effects on age at vaginal opening; in two of these studies, effects on body weight/body weight gain also were noted. Two of these studies also detected significant differences in body weight at vaginal opening. In general, statistical significance in adjusted values for age of vaginal opening, body weight at vaginal opening, and organ weights correlated well with statistical significance in absolute values for these parameters in the female pubertal assay. Two compounds altered percent regularly cycling, but only in the presence of a significant change in BUN or body weight changes. One compound induced significant changes in uterine weights (both wet and blotted), but did not affect any other estrogen-sensitive endpoints. For this study, six high-dose females were in proestrus or late diestrus on the day of necropsy compared to one control group female in late diestrus on the day of necropsy, suggesting that the higher uterine weights in the absence of effects on estrogen-sensitive endpoints were the result of the stage of the estrous cycle on the day of necropsy and not indicative of an estrogenic response. Adjusting values based on body weight at the start of dosing did not change the interpretation of any of these studies. Two compounds affected ovarian and uterine histopathology; one was an antiestrogen and the other compound produced an apparent developmental delay.

When examining thyroid endpoints across both the male and female pubertal assays, there were five compounds that produced statistically significant changes in thyroid histopathology using the five-point grading scale, but eight compounds that caused qualitative changes in thyroid histopathology as judged by pathologists. Of these eight compounds, five caused significant changes in serum T4 and/or TSH levels and three caused changes in thyroid weight. Two compounds caused thyroid histopathological changes without any associated

changes in other thyroid endpoints (thyroid hormones or weight). Five compounds had significant changes in T4 and/or TSH without any associated changes in thyroid weight or histopathology; in each of these cases, there were effects on body weight/body weight gain.

Some examples shown in Tables 8 and 9 lack concordant changes across related, endocrine-sensitive endpoints; thus clear interpretation of pubertal assay results or MoA may be problematic. One of the difficulties when using the pubertal assays is the reliance of these assays on apical endpoints. As a result, statistically significant effects observed in the pubertal assays may not be diagnostic of endocrine activity. The redundancy of the entire Tier 1 screening battery should aid in interpretation of pubertal assay results since a WoE approach will be used to determine whether effects in Tier 1 warrant further evaluation in Tier 2 tests; however, there is still concern over how this WoE approach will be implemented. For example, one of the more problematic endpoints in the pubertal assays is the attainment of puberty (i.e., the age and body weight at which the animals achieve vaginal patency for females or PPS for males). It is well documented that numerous factors can impact these endpoints (e.g., Frisch et al., 1975; Zipf et al., 1978; Smith et al., 1989; Cicero et al., 1990, 1991; Odum et al., 2001). However, it is still unclear how EPA will view a statistically significant effect on these endpoints, particularly if the difference is within the range of normal variance and how such effects may influence the overall WoE evaluation and triggers for Tier 2 testing. A hypothesis-driven WoE approach has been proposed by Borgert et al. (2011b) for testing the premise that a substance interacts as an agonist or antagonist with components of estrogen, androgen, or thyroid pathways or with components of the aromatase or steroidogenic enzyme systems. This approach proposes deriving response and relevance weightings based on the assay results and provides a framework for use in performing a WoE assessment.

One of the benefits of *in vivo* models is the ability to add additional mechanistic endpoints to aid in interpretation. Expanding the number of hormones included in the analyses can greatly enhance the MoA understanding of the effects, and strengthen understanding of the apical endpoints that may be affected. This has been done previously in another screening battery using adult animals (O'Connor et al., 2005), where the utility of a comprehensive hormonal assessment was shown to greatly enhance the ability to identify MoA for a series of positive control compounds. Since limited hormonal analyses are already included as part of the pubertal assays, this could easily be expanded by collecting a greater volume of blood at necropsy. Addition of other endpoints also should be considered if the nature of the test substance warrants it. For example, it may be useful to collect and save livers for possible biochemical evaluation if triggered by other findings. This adds very little cost, but allows flexibility to run additional endpoints to help further define MoA. A good example is evaluating hepatic UDP-glucuronyltransferase as a means to look at effects on thyroid hormone perturbations (O'Connor et al., 2002). Alternatively, if a substance induces liver enzymes, resulting in enhanced clearance of steroid hormones, these liver samples can further characterize this substance as one that may

cause effects on endocrine parameters, but secondary to effects on the liver. At a minimum, all tissues that are weighed should be saved for possible histopathological evaluation if findings warrant it. In addition, other tissues could be saved as a precaution, for example, adrenals, pituitaries, etc. While the recommendation would not include routine evaluation of these tissues, if they are saved at the time of necropsy, they would be available for evaluation if findings warranted analysis. The benefit of including additional endpoints must be weighed against the cost of including them. At a minimum, the recommendation would be to save all weighed tissues for possible histopathological evaluation, collect a larger volume of blood, and preserve serum at -70°C for possible hormonal analyses, and to flash freeze and save a portion of the liver at -70°C for possible biochemical analyses.

PERFORMANCE CRITERIA

The U.S. EPA test guidelines for the male and female pubertal assays include performance criteria for control data, which laboratories should strive to meet to confirm that the pubertal assays were conducted properly. The test guidelines state that "mean values and coefficients of variation (CVs) for the vehicle control group must fall in the acceptable range of each to be considered fully acceptable." Notably, not all endpoints have specified values for performance criteria; in Sprague-Dawley rats, there are no performance criteria for mean range for testis weights, and mean range and CVs for TSH levels and weaning body weight in females (other criterion are missing in Wistar rats). Furthermore, data on hormone levels can vary with measurement methodology (Soldin et al., 2004; Hegstad-Davies, 2006) and the methods used to generate the performance criteria for hormone data were not specified in the test guidelines.

During the pubertal assay validation program, interlaboratory studies showed that these performance criteria were difficult to meet. None of the laboratories met all of the mean range and CV performance criteria in either the male or female pubertal assays (U.S. EPA, 2007a, 2007b). In the male pubertal assay, none of the three laboratories had mean control values within the acceptable range for all endpoints and none met acceptable CV values for all endpoints in control animals. In the female pubertal assay, only one of three laboratories had mean control values within the acceptable range for all endpoints and none of the laboratories met acceptable CV values for all endpoints in control animals.

These interlaboratory validation results were similar to the results reported by laboratories contributing data to the current article. In the male pubertal assay, the recommended mean control values for kidney and thyroid weights were below the lowest recommended value in 100 and 67% of studies, respectively (Table 10). Single studies missed the mean control ranges for body weight at weaning, LABC weight, and serum testosterone levels. Across endpoints, maximum CV values were exceeded in 0–29% of the male pubertal assays conducted in our laboratories with higher rates of exceedance observed with age at PPS (38% of studies), body weight at PPS (57% of studies), final body weight (57% of studies), and weight of the seminal vesicles with coagulating glands

Table 10
Performance Criteria for Male Pubertal Interlaboratory Control Values¹

Endpoint	Recommended mean range	No. (%) of studies outside mean range	Recommended maximum CV	No. (%) of studies where CV was exceeded
Age at PPS ^a	39.781–46.513	0 (0%)	5.67	3 (14%)
Body wt at PPS (g)	188.277–256.169	0 (0%)	7.57	17 (81%)
Body wt at weaning (g)	45.472–59.812	1 ^b (5%)	10.25	3 (14%)
Final body wt (g)	259.235–332.059	0 (0%)	7.47	15 (71%)
Liver wt (g)	9.990–15.350	0 (0%)	14.93	2 (10%)
Kidney wt (g)	2.242–3.050	21 ^b (100%)	14.76	1 (5%)
Pituitary wt (mg)	7.810–12.898	0 (0%)	15.98	7 (33%)
Ventral prostate wt (g)	0.160–0.332	0 (0%)	22.32	3 (14%)
SV wt w/CG with fluid (g)	0.295–0.719	0 (0%)	21.06	8 (38%)
LABC wt (g)	0.447–0.855	1 ^b (5%)	27.10	0 (0%)
Thyroid wt (mg)	14–26	14 ^b (67%)	23.63	4 (19%)
Serum TSH levels (ng/ml)	4.212–24.112	0 (0%)	58.29	6 (29%)
Serum T levels (ng/ml)	0.260–3.960	1 ^c (5%)	89.70	0 (0%)

n = 21 studies.

PPS, preputial separation; wt, weight; SV w/CG, seminal vesicles with coagulating glands; T, testosterone.

¹Values for adrenal weight, epididymis weight, and serum T4 levels met both mean range and CV performance criteria.

^aPostnatal day (PND), where day of birth = PND 0.

^bMean values were lower than recommended values.

^cMean values were higher than recommended values.

and fluid (38% of studies). Again, the variables for which the maximum recommended CV was exceeded by our laboratories were generally the same variables that failed to meet these criteria during the U.S. EPA's interlaboratory validation. In the interlaboratory validation, two of three laboratories exceeded the maximum CV values for age at PPS, final body weight, pituitary weight, weight of the seminal vesicles with coagulating glands and fluid, and thyroid weight; all three laboratories exceeded the maximum CV value for body weight at PPS and ventral prostate weight (U.S. EPA, 2007a). The mean control values and CVs for the remaining assay variables (i.e., serum T4 levels, adrenal, and epididymal weights) met the performance criteria in our laboratories.

In the female pubertal assay, our laboratories generally met the recommended range for mean control values, although 87% of studies had mean control adrenal weights that were lower than the recommended range (<38.34 mg) and 26% of studies had mean control values for age at vaginal opening that were higher than the maximum performance criteria value (>35.62 days of age) (Table 11). One study missed the performance criteria for mean body weight at vaginal opening. The maximum recommended CV values were more problematic as maximum control CV values generally were exceeded for five variables in 9–26% of studies (Table 11). The variables for which the maximum recommended CV was exceeded by our laboratories were generally the same variables that failed to meet these criteria during the U.S. EPA's interlaboratory validation. In the interlaboratory validation, two of three laboratories exceeded the maximum CV values for final body weight, age at vaginal opening, and kidney weights; all three laboratories exceeded the maximum CV value for liver weight (U.S. EPA, 2007a). In addition to the parameters for which the maximum CV was exceeded in the interlaboratory validation, body weight at

vaginal opening was exceeded in two of the studies performed by the authors' laboratories (Table 11). The mean control values and CVs for the remaining assay variables (i.e., uterine weight, ovarian weight, T4, thyroid weight, and pituitary weight) met the performance criteria in our laboratories.

Thus, based on the collective experience of the authors' laboratories, as well as the results from the interlaboratory validation of these assays, we suggest that the EPA may want to adjust some of the performance criteria for the pubertal assays. The frequency with which the same performance criteria were repeatedly missed by multiple laboratories was specifically noted by the FIFRA Scientific Advisory Panel, which reviewed the EDSP Tier 1 assays in May 2013. The Scientific Advisory Panel stated, "The Agency (EPA) should look at adjusting the performance criteria for the individual assays to better reflect the experience gained since 2008" (FIFRA SAP, 2013).

RECOMMENDATIONS

One of the greatest challenges for the male and female pubertal assay is dose selection. As shown in this article, numerous endpoints have been used to define the MTD. A careful review of previous toxicity data, coupled with range-finding studies in juvenile animals, may be needed to set dose levels that achieve an MTD, yet minimize the potential for systemic toxicity that can cloud assay interpretation.

For substances that are not potent hormonal agents, it will likely be difficult to determine whether effects in the male and female pubertal assays are diagnostic of a primary, endocrine-mediated change or whether effects are secondary to systemic toxicity, body weight changes, stress, stage of estrous, etc. If effects are judged to be specifically endocrine mediated, discerning the MoA may

Table 11
Performance Criteria for Female Pubertal Interlaboratory Control Values¹

Endpoint	Recommended mean range	No. (%) of studies outside mean range	Recommended maximum CV	No. (%) of studies where CV was exceeded
Final body wt (g)	104.86–204.55	0 (0%)	8.93	4 (17%)
Age at VO ^a	30.67–35.62	7 ^b (30%)	6.52	5 (22%)
Body wt at VO (g)	101.71–131.44	1 ^b (4%)	13.97	2 (9%)
Liver wt (g)	4.32–11.78	0 (0%)	13.13	6 (26%)
Kidney wt (g)	0.95–2.20	0 (0%)	10.76	2 (9%)
Adrenal wt (mg)	38.34–48.84	20 (87%)	22.97	0 (0%)

n = 23 studies.

wt, Weight; VO, vaginal opening.

¹Values for pituitary weight, ovarian weight, blotted uterine weight, thyroid weight, and serum T4 levels met both mean range and CV performance criteria.

^aPostnatal day (PND), where day of birth = PND 0.

^bMean values were higher than recommended values.

be difficult. Patterns of effects should be apparent if specific endocrine effects have occurred, although these patterns alone are not diagnostic. To aid in assay interpretation, additional endpoints may be included (e.g., preserve tissues for additional histopathology, collect greater serum volumes to analyze additional hormones, and save frozen livers to examine enzyme induction). Of the utmost importance, the pubertal assays should be interpreted as part of the EDSP battery to take advantage of assay redundancy to detect endocrine MoAs.

Based on the frequency with which certain performance criteria were missed, an EPA review of these criteria is warranted. This recommendation was supported by the FIFRA Scientific Advisory Panel in their May 2013 review of the EDSP Tier 1 battery (FIFRA SAP, 2013).

Lastly, it would be beneficial if the U.S. EPA gave additional guidance as to how performance criteria will be used and what determines assay acceptability. It would be helpful for registrants to understand these factors to determine whether a study should be repeated. This would avoid a delay for both the test order recipients and the EPA, which may find data unsuitable at the time of review and thus, delay its decision making.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the following individuals for assistance during these studies: A. Andrus, C. Zablony, J. Passage, K. Gallagher, M. Lawson, J. Thomas, R. Sura, R. Hukkanen K. Stebbins, P. Sawhney Coder, E. Slotter, J. Toot, M. Herberth, and the TERC and WIL Research pathology and animal care groups. The authors also wish to recognize the contributions of Dr. Robert Ellis-Hutchings in reviewing this article.

CONFLICT OF INTEREST

These studies were conducted to meet data call-in requirements as part of the U.S. EPA EDSP, and therefore,

the studies were funded by companies that received test orders for this regulatory program.

REFERENCES

- Ashby J, Lefevre PA. 2000. The peripubertal male rat assay as an alternative to the Hershberger castrated male rat assay for the detection of antiandrogens, oestrogens and metabolic modulators. *J Appl Toxicol* 20:35–47.
- Boelen A, Wiersinga WM, Fliers E. 2008. Fast induced changes in the hypothalamus-pituitary-thyroid axis. In: *Thyroid economy regulation, cell biology, thyroid hormone metabolism and action: The special edition: Metabolic effects of thyroid hormones*. *Thyroid* 18:123–129.
- Borgert CJ, Mihaich EM, Quill TF, Marty MS, Levine SL, Becker RA. 2011a. 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, Mihaich EM, Ortego LS, Bentley KS, Holmes CM, Levine SL, Becker RA. 2011b. Hypothesis-driven weight of evidence framework for evaluating data within the U.S. EPA's Endocrine Disruptor Screening Program. *Regul Toxicol Pharmacol* 61:185–191.
- Burkhardt CA, Robinson JL. 1978. High rat pup mortality attributed to the use of cedar-wood shavings as bedding. *Lab. Anim.* 12:221–222.
- Carney EW, Zablony CL, Marty MS, Crissman J, Anderson P, Woolhiser M, Holsapple M. 2004. The effects of feed restriction during *in utero* and postnatal development in CD rats. *Toxicol Sci* 82:237–249.
- Cicero TJ, Adams ML, O'Connor L, Nock B, Meyer ER, Wozniak D. 1990. Influence of chronic alcohol administration on representative indices of puberty and sexual maturation in male rats and the development of their progeny. *J Pharmacol Exp Ther* 255:707–715.
- Cicero TJ, Adams ML, Giordano A, Miller BT, O'Connor L, Nock B. 1991. Influence of morphine exposure during adolescence on the sexual maturation of male rats and the development of their offspring. *J Pharmacol Exp Ther* 256:1086–1093.
- Conolly RB, Beck BD, Goodman JI. 1999. Stimulating research to improve the scientific basis of risk assessment. *Toxicol Sci* 49:1–4.
- Döhler K-D, Wong CC, von zur Muhlen A. 1979. The rat as model for the study of drug effects on thyroid function: consideration of methodological problems. *Pharmacol. Ther.* B 5:305–318.
- Eales JG. 1988. The influence of nutritional state on thyroid function in various vertebrates. *Amer. Zool.* 28:351–362.
- FIFRA SAP. 2013. Transmittal of meeting minutes from the FIFRA Scientific Advisory Panel held May 21–23, 2013 on "Endocrine Disruptor Screening Program (EDSP) Tier 1 Screening Assays and Battery Performance". Memorandum to the Office of Science Coordination and Policy and the Office of Pesticide Programs, US EPA, dated August 21, 2013. Available at: <http://www.epa.gov/scipoly/sap/meetings/2013/may/052113minutes.pdf>. Accessed September 24, 2013.

- Frisch RE, Hegsted DM, Yoshinaga K. 1975. Body weight and food intake at early estrus of rats on a high-fat diet. *Proc Natl Acad Sci USA* 72:4172–4176.
- Hegstad-Davies RL. 2006. A review of sample handling considerations for reproductive and thyroid hormone measurement in serum or plasma. *Theriogenology* 66:592–598.
- Juberg DR, Borghoff S, Becker RA, Casey W, Hartung T, Holsapple M, Marty S, Mihaich E, Van Der Kraak G, Wade MG, Willett K, Andersen M, Borgert C, Coady K, Dourson M, Fowle JR III, Gray E, Lamb J, Ortego L, Schug TT, Toole C, Zorrilla L, Kroner O, Patterson J, Rinckel L, Jones B. ^t Workshop report: lessons learned, challenges, and opportunities: the U.S. Endocrine Disruptor Screening Program. *Altex* 31:63–78 (Jan. 2014). Available at: <http://www.altex.ch/resources/WRJuberg.epub.pdf> pii:S1868696x1309271X. Accessed October 27, 2013.
- Laws SC, Ferrell JM, Stoker TE, Schmid J, Cooper RL. 2000. The effect of atrazine on female wistar rats: an evaluation of the protocol for assessing pubertal development and thyroid function. *Toxicol Sci* 58:366–376.
- 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.
- Markaverich BM, Alejandro MS, Markaverich D, Zitzow L, Casajuna N, Camarao N, Hill J, Bhirdo K. 2002. Identification of an endocrine disrupting agent from corn with mitogenic activity. *Biochem Biophys Res Commun* 291:692–700.
- Marty MS, Crissman JW, Carney EW. 1999. Evaluation of the EDSTAC female pubertal assay in CD rats using 17 β -estradiol, steroid biosynthesis inhibitors, and a thyroid inhibitor. *Toxicol Sci* 52:269–277.
- Marty MS, Johnson KA, Carney EW. 2003. Effect of feed restriction on Hershberger and pubertal male assay endpoints. *Birth Defects Res B* 68:363–374.
- Matysek M. 1989. Studies on the effect of stress on the estrus cycle in rats. *Ann Univ Mariae Curie Sklodowska Med* 44:143–149.
- O'Connor JC. 2005. The 15-day intact adult male as an alternative Tier 1 screening assay for detecting endocrine active compounds. Slide presentation at: http://www.epa.gov/scipoly/oscpendo/presentations/intact_male.presentation.080205.ppt. Accessed September 24, 2013.
- O'Connor JC, Frame SR, Davis LG, Cook JC. 1999a. Detection of the environmental antiandrogen p,p'-DDE in CD and Long-Evans rats using a Tier I screening battery and a Hershberger assay. *Toxicol Sci* 51:44–53.
- O'Connor JC, Frame SR, Davis LG, Cook JC. 1999b. Detection of thyroid toxicants in a Tier 1 screening battery and alterations in thyroid endpoints over 28-days of exposure. *Toxicol Sci* 51:54–70.
- O'Connor JC, Frame SR, Ladics GS. 2002. Evaluation of a 15-day screening assay using intact male rats for identifying steroid biosynthesis inhibitors and thyroid modulators. *Toxicol Sci* 69:79–91.
- Odum J, Tinwell H, Jones K, Van Miller JP, Joiner RL, Tobin G, Kawasaki H, Ashby J. 2001. Effect of rodent diets on the sexual development of the rat. *Toxicol Sci* 61:115–127.
- Organisation for Economic Co-Operation and Development (OECD). 2007. *Guidelines for Testing of Chemicals*, Section 4: *Health Effects*, protocol number 440, uterotrophic bioassay in rodents: a short-term screening test for oestrogenic properties <http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents.9789264067417-en;jsessionid=12ic71yg7bopl.delta>. Accessed January 29, 2014.
- Rooszendaal MM, Swarts HJ, Wiegant VM, Mattheij JA. 1995. Effect of restraint stress on the preovulatory luteinizing hormone profile and ovulation in the rat. *Eur J Endocrinol* 133:347–353.
- Smith BJ, Plowchalk DR, Sipes IG, Mattison DR. 1991. Comparison of random and serial sections in assessment of ovarian toxicity. *Reprod Toxicol* 5:379–383, as cited in Plowchalk DR, Smith BJ, Mattison DR. "Assessment of toxicity to the ovary using follicle quantitation and morphometrics". Chapter 5 in Tyson CA, Witschi H. *Methods in Toxicology*, Vol. 3B, Female reproductive toxicology, Heindel JJ, Chapin RE, eds., 1993.
- Smith SS, Neuringer M, Ojeda SR. 1989. Essential fatty acid deficiency delays the onset of puberty in the female rat. *Endocrinology* 125:1650–1659.
- Soldin OP, Tractenberg RE, Soldin SJ. 2004. Differences between measurements of T4 and T3 in pregnant and nonpregnant women using isotope dilution tandem mass spectrometry and immunoassays: are there clinical implications? *Clin Chim Acta* 347:61–69.
- Stoker TE, Laws SC, Guidici DL, Cooper RL. 2000. The effect of atrazine on puberty in male wistar rats: an evaluation in the protocol for the assessment of pubertal development and thyroid function. *Toxicol Sci* 58:50–59.
- U.S. EPA. 2007a. Integrated summary report for validation of a test method for assessment of pubertal developmental and thyroid function in juvenile female rats as a potential screen in the Endocrine Disruptor Screening Program Tier-1 Battery. Available at: http://www.epa.gov/endo/pubs/female_isr.v4.1c.pdf. Accessed January 29, 2012.
- 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. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/male_pubertal_isr.pdf. Accessed January 29, 2012.
- U.S. EPA. 2009a. Series 890-Endocrine Disruptor Screening Test OPPTS 890.1450: pubertal development and thyroid function in intact juvenile/peripubertal female rats. Available at: <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0009>. Accessed January 29, 2012.
- U.S. EPA. 2009b. 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 January 29, 2012.
- U.S. EPA. 2011a. United States Environmental Protection Agency. Weight-of-evidence: evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. Document ID: 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. 2011b. Memorandum: "Corrections and Clarifications on Technical Aspects of the Test Guidelines for the Endocrine Disruptor Screening Program Tier 1 Assays (OCSP Test Guideline Series 890)," (March 4, 2011). Available at: <http://www.epa.gov/scipoly/oscpendo/pubs/toresources/clarificationdoc.pdf>. Accessed September 25, 2013.
- U.S. EPA. 2013. Endocrine Disruptor Screening Program Tier 1 assays: considerations for use in human health and ecological risk assessments (June 2013). Available at: http://www.epa.gov/scipoly/oscpendo/pubs/use_of_tier_1.data.in.risk.assessment.pdf. Accessed September 8, 2013.
- You L, Casanova M, Bartolucci EJ, Fryczynski MW, Dorman DC, Everitt JJ, Gaido KW, Ross SM, Heck HD. 2002. Combined effects of dietary phytoestrogen and synthetic endocrine-active compound on reproductive development in Sprague-Dawley rats: genistein and methoxychlor. *Toxicol Sci* 66:91–104.
- Zarrow MX, Yochim JM, McCarthy JL, Sanborn RC. 1964. Experimental endocrinology. A sourcebook of basic techniques. New York: Academic Press.
- Zipf WB, Payne SH, Kelch RP. 1978. Prolactin, growth hormone, and luteinizing hormone in the maintenance of testicular luteinizing hormone receptors. *Endocrinology* 103:595–600.