Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats Assay

OCSPP Guideline 890.1500

Standard Evaluation Procedure (SEP)

ENDOCRINE DISRUPTOR SCREENING PROGRAM
U.S. Environmental Protection Agency
Washington, DC 20460

August 2011

Table of Contents

I. INTRODUCTION 1

A. Use of the Standard Evaluation Procedure 1

II. THE MALE PUBERTAL ASSAY 1

A. Purpose of the Assay 1

B. Background 2

C. Study Design 2

D. Endpoints Evaluated 2

III. EVALUATION OF STUDY CONDUCT 3

A. Test Compound 3

B. Test Animals 3

C. Animal Husbandry 4

D. Experimental Design 5

E. Dosing 5

F. Preputial Separation 5

G. Terminal Procedures 6

1. Necropsy 6

2. Hormone Assays 8

3. Blood Chemistry 8

4. Histology 8

H. Statistical Analyses 9

IV. STUDY INTERPRETATION 10

A. Results 10

1. General Growth and Preputial Separation 10

2. Organ Weights at Necropsy 12

3. Thyroxine, Thyroid Stimulating Hormone, and Testosterone Levels, and Clinical Chemistry 14

4. Histopathology 14

B. Data Interpretation Procedure 16

C. Performance Criteria 17

V. CHARACTERIZATION OF FINDINGS 19

VI. DATA EVALUATION RECORD 19

VII. REFERENCES 19

I. INTRODUCTION

A. Use of the Standard Evaluation Procedure

This Standard Evaluation Procedure (SEP) was developed by the U.S. Environmental Protection Agency (EPA) to provide guidance to EPA staff who will be reviewing the data submitted in response to Tier 1 Orders issued under the Endocrine Disruptor Screening Program (EDSP). The SEPs provide general guidance and are not binding on either EPA or any outside parties. The use of language such as "will," "is," "may," "can" or "should" in these documents does not connote any requirement for either EPA or any outside parties. As such, EPA may depart from the guidance where circumstances warrant and without prior notice. The SEPs are intended to be used in conjunction with the EDSP Test Guideline Series 890 and the Corrections and Clarifications document available on the EDSP web page.

This SEP provides guidance on how to review studies conducted using the OCSPP Guideline 890.1500 for the Male Pubertal Assay that are submitted to support requirements imposed under the EPA's EDSP. The product of the review will be a Data Evaluation Record (DER) that reflects how well the study was performed and conforms to the Guideline; and provides the appropriate conclusions supported by the data. The DER will include, for example, a list of any significant deviations from the protocol as well as their potential impacts, a list of significant information missing from the study report, and any other information about the performance of the study that affects interpretation of the data within the context of the EDSP.

The DER should contain adequate information to provide the EPA with the ability to determine whether the study was performed according to the Guideline. The objective of EDSP Tier 1 assays is to characterize the potential of a chemical to interact with the endocrine system.

The Guideline recommends the critical materials, methods, and analyses that lead to successful performance of the assay. If a particular material, method, or analysis is named in the Guideline, it is usually because other materials, methods, or analyses are either known to be inappropriate—or at least have not been validated—or that there is concern for their potential influence on results. The Agency has posted Corrections and Clarifications on Technical Aspects of the EDSP Tier 1 Assays (OCSPP Test Guideline Series 890) in the docket; the link to this document may be found by way of the EDSP web page (http://www.epa.gov/endo/). It is therefore important to note deviations from specific materials, methods, or analyses in the DER, and provide the Agency's opinion on whether the deviation/deficiency has an impact on the performance and results of the study or the acceptability of the study.

II. THE MALE PUBERTAL ASSAY

A. Purpose of the Assay

The purpose of the male pubertal assay is to provide information obtained from an *in vivo* mammalian system that will be useful in assessing the potential of a chemical substance or mixture to interact with the endocrine system. This assay is capable of detecting not only anti-thyroid, androgenic, or anti-androgenic chemicals, but also agents that alter pubertal development through mechanisms that induce changes in gonadotropins, prolactin or via alterations in hypothalamic function.

B. Background

Endocrine disrupting xenobiotics have the potential to interact with hormone systems. The EDSP reflects a two-tiered approach to implement the statutory testing requirements of FFDCA section 408(p) (21 U.S.C. 346a) and SDWA (42 U.S.C. 300j-17). In general, EPA intends to use the data collected under the EDSP, along with other scientifically relevant information (OSRI), to determine if a pesticide, chemical, or other substance may pose a risk to human health or the environment due to disruption of the endocrine system. EPA has developed the OCSPP 890 Series guidelines as a screening battery (Tier 1 screening) to identify these substances.

C. Study Design

In the general design of a Pubertal Male Rat study, groups of juvenile male rats are exposed to the test substance daily by oral gavage from post-natal day (PND) 23 through PND 53. This duration of treatment is required for the detection of pubertal delay and anti-thyroid effects. The rats are either treated with vehicle only, or the test substance in the vehicle at two dose levels. The evaluation includes observations to detect thyroid changes, androgenic or anti-androgenic chemicals or agents that alter pubertal development.

D. Endpoints Evaluated

The following endpoints are evaluated in the Male Pubertal Assay:

Growth	Daily body weight
Preputial Separation (PPS)	Age and weight at PPS
Organ weights	Seminal vesicle plus coagulating glands (with and without fluid) Ventral prostate Dorsolateral prostate Levator ani/bulbocavernosus (LABC) muscle complex Epididymides (left and right separately) Testes (left and right separately) Thyroid (after fixation) Liver Kidneys (paired) Pituitary Adrenals (paired)
Histology	Epididymis (one) Testis (one) Thyroid (colloid area and follicular cell height) Kidney
Hormones	Serum testosterone, total Serum thyroxine (T_4), total Serum thyroid stimulating hormone (TSH)
Clinical (serum) chemistry	Standard blood panel, including creatinine and blood urea nitrogen (BUN)

III. EVALUATION OF STUDY CONDUCT

A. Test Compound

It is important to report the purity of the test compound used in the study in the DER, along with its source, Lot No. and/or Batch No., and whether the certificate of analysis was provided. The stability of the test chemical in the vehicle is also a relevant point to be reported in the DER for concentrations that bracket those used in the study, along with the storage temperature used for the stability analyses. For test substances dosed as suspensions, the homogeneity of the test chemical suspensions is also relevant information to be reported. EPA recommends that the DER provide the range of values expressed as percent of nominal for the analysis of the test chemical solutions.

EPA recommends that the dosing vehicle used in the study not have any potential intrinsic toxicity (e.g., acetone, DMSO) in order to avoid confounding the study results. Other solvents, such as water or carboxymethylcellulose, may also be appropriate. Use of an intermediate solvent would not be expected to affect the validity of the data significantly, provided the concentration is kept at low concentrations (e.g., 1% or below) and is used across all test groups including control. EPA recommends that the study design include the use of a solvent or vehicle control group.

Gentle warming may be effectively used to assist solubilization, but it is recommended that the solution not be administered warm and be checked to make sure that precipitation did not occur upon cooling. If the test substance is not soluble in any of the conventional solvents, EPA recommends that it be administered as a suspension. It is important that the dosing solution or suspension be well-mixed to keep the chemical well-distributed prior to and throughout dosing, and care must be taken to ensure that the particle size of insoluble substances does not interfere with delivery of the full dose through the gavage tube or needle tip.

B. Test Animals

Sprague-Dawley rats are the preferred strain for this assay until a more-appropriate strain (or set of strains) is identified and associated performance criteria are developed. Results similar to those from Sprague-Dawley rats have been produced using Wistar and Long-Evans rats in this assay or relevant modifications of this assay, suggesting that strain is not the major determinant of sensitivity in this assay. EPA recommends that the DER include: the species and strain used and the rationale for this choice; the source and supplier of the animals; and the number and age of the animals at receipt, dosing initiation, and necropsy.

It is recommended that juvenile male rats be derived from individually housed pregnant females that were either bred in-house or purchased from a supplier as “timed pregnant” dams. EPA recommends that dams obtained and transported from an external supplier not be used in the same study as dams bred in-house. There are several reasons that the Agency strongly recommends that pregnant dams rather than pups of a specified age be used as the starting point for the pubertal assays. First, it is important to minimize genetic effects (that is, litter effects) on the endpoints of interest where possible. It is important to know which pups are from the same dam in order to be able to randomize distribution of siblings across treatment groups. Second, it

is important to cull to 8 to 10 pups per litter within 3 or 4 days after birth. This helps to minimize variability in body weights across a litter, and thus minimize variability in day or preputial separation and in other endpoints such as organ weights that may be related to body weight. Third, it is important to know the day of birth of each pup accurately. Knowing the day of birth accurately is critical for accurate determination of age at preputial separation and keeping the coefficients of variation (CVs) low. Keeping the CVs as low as possible is important for maximizing the sensitivity of the pubertal endpoints.

Thus, if pups of a specified age are ordered rather than pregnant dams, the Agency recommends that submitters document that all of the necessary steps have still been taken to allow randomization of litters across treatment groups, and that standardization of litters (including exclusion of litters with fewer than 8 total pups per litter as well as litters not delivered by GD 23, reduction of litter size to 8-10 pups per litter between PND 3 and 5, and not allowing cross-fostering) has been done.

It is also recommended that all dams be pregnant for the first time and timed to deliver on the same day. If purchased from a supplier, EPA recommends that all dams be on the same gestation day (GD), but does not believe that whether the day is GD 7, 8, 9 or 10 at the time of arrival at the performing laboratory (where GD 0 is defined as the day the dams are found to be sperm positive) is likely to affect study results. EPA recommends that dams be allowed to deliver their pups naturally.

The Guideline recommends pup body weights be monitored weekly and any unthrifty litters or runted pups not be included in the study. The Guideline also recommends littermates not be used in the same experimental group; and at least 15 male pups be used per treatment group. Typically, the pups are weaned on PND 21, and the male pups assigned to treatment groups such that the mean body weights and variances for all groups are similar.

C. Animal Husbandry

The Guideline recommends that the study be conducted in facilities accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC) if in the U.S., or the applicable national or international accreditation authority if outside the U.S. Rats are housed in cages with heat-treated (to eliminate resins that induce liver enzymes) laboratory-grade wood shavings (other than cedar) as bedding. The Guideline recommends corn cob bedding not be used due to its potential to disrupt endocrine activity (Markaverich *et al*., 2002); and wire-mesh-bottomed caging not be used due to the potential for pup loss.

The Guideline recommends animals be maintained on a balanced laboratory diet that has a genistein-equivalent content of genistein plus daidzein (aglycone forms) less than or equal to 300 µg/g, and the same batch of feed be used for treated and control groups at all times.

Deionized water is recommended in the Test Guideline. Other acceptable sources of water include double-distilled water and charcoal-filtered water. Other sources may be acceptable; however, the presence of soluble organic chemical contaminants such as natural or artificial hormones have the potential to introduce variability into (and potentially compromise) the results. If an alternative source of water has been used, the Agency recommends that the

laboratory document that such contaminants have been removed from the drinking water. The Test Guideline recommends that water not be supplied using polycarbonate supply equipment.

D. Experimental Design

The Guideline specifies use of a randomized complete block design (time-separated necropsy is the blocking factor) with at least 15 male rats in each treatment group. The treatment groups include: (1) the vehicle-treated and (2) xenobiotic-treated with at least two dose levels. The Guideline recommends that the highest dose level be at or just below the Maximum Tolerated Dose (MTD) but need not exceed the Limit Dose (1000 mg/kg/day); however, typically, the Agency also considers the toxicity profile of the chemical (i.e., cholinesterase inhibition, target organ toxicity, etc) in dose selection. EPA recommends that the second dose level typically be spaced to produce a lesser degree of toxicity relative to the high dose unless justification is provided for testing at a different level. The DER should contain the rationale provided for the selection of doses. If necessary, it is possible to conduct the study in time-separated blocks rather than at one time without affecting the validity of the results. In this case, each block is recommended to contain all treatment groups (i.e., control, vehicle control, etc.) and be balanced with respect to numbers of animals and body weight at weaning.

E. Dosing

The Guideline recommends that animals be weighed daily, prior to treatment, and the body weight recorded. Consistent with good scientific practice, clinical observations are recorded daily. Endpoint measurements (organ weights, hormone levels, histology, etc.) from animals which were found dead or euthanized *in extremis* during the study are not included in the summary statistics.

The Guideline specifies treatments be administered daily by oral gavage from PND 23 through PND 53, in order to detect delays in puberty and anti-thyroid effects. The Agency recommends test chemicals be administered daily by gavage at a dose volume of 2.5 to 5.0 mL vehicle/kg body weight at 0700-0900 daily. The Guideline recommends dosing equipment not utilize any plastic or rubber to avoid the potential for absorption by or leaching of substances. The treatments are typically administered on a mg/kg body weight basis, using the current day's weight, and volume of the dose administered recorded each day.

In the absence of other clinical signs that would normally lead to removal of an animal from the study, failure to gain weight at the same rate as controls is generally considered not to be a reason to remove a treated animal during the course of the study. However, it is recognized that severe failure to grow may be a reason to disqualify an animal even in the absence of other signs of toxicity. These types of clinical sign are to be reported in the DER.

F. Preputial Separation

Beginning on PND 30, males are examined daily for preputial separation (PPS). The appearance of partial and complete PPS, or a persistent thread of tissue between the glans and prepuce are all recorded on the days they are observed. The day of complete preputial separation is the endpoint used in the analysis for the age at PPS. However, if any animal within any treatment group shows incomplete separation (including persistent threads) for greater than three days, a separate analysis is typically conducted using the ages at which partial separation was

first observed. Documentation of a thread, even if PPS otherwise appears complete, is important. It is also critical that "initiation" of preputial separation be recorded. However, some animals may complete separation within a day and the initiation may not be observed. It is preferred, but not critical, that PPS observations be taken after the daily dosing. Whether collected before or after dosing, it is critical that the preputial separation observations be collected at approximately the same time each day in order to detect accurate timing of initiation of PPS. The age and body weight at PPS is generally reported in the DER for each treatment group.

It is critical that, for each animal, the PPS observation be recorded for the day immediately prior to the day on which PPS begins. Laboratories may choose not to begin monitoring on PND 30, but missing the day on which PPS begins for each animal will be considered a serious deficiency in the study because the sensitivity of this endpoint is dependent on the accurate determination of the day of PPS (or in certain cases, day of initiation of PPS if the process is not complete in one day). If the day of PPS is expected to be different for control animals from what is noted in the Guideline, EPA recommends that appropriate documentation to support a modification in study design to begin observations later than recommended be provided to the Agency. Note that consideration should be given to the possibility that the test chemical may accelerate PPS; consequently, observations for these endpoints should begin substantially before the age at which control animals are expected to reach these endpoints. The objective of the PPS endpoint is to determine quantitatively the difference, if any, in age at PPS between treated groups and controls, not merely to determine that an acceleration or delay has occurred.

G. Terminal Procedures

1. Necropsy

The Guideline recommends males be sacrificed on PND 53 beginning 2 hours after the final dose. If necessary, one half of the males can be sacrificed on PND 53 and the remaining males on PND 54 as long as the animals in each treatment group (including controls) are equally dispersed between the two necropsy days. The Test Guideline recommends that animals sacrificed on PND 54 be dosed and treated on the day of sacrifice just like the animals sacrificed on PND 53 with regard to time of dosing, collection of preputial separation, etc. It is critical that sacrifices are completed by 1300 hours due to normal diurnal fluctuation in thyroid hormone levels. In order to minimize animal stress, the Guideline recommends sacrifices be performed in a separate room from the holding room and that the time from transfer between rooms to sacrifice be as brief as possible.

The preferred method of sacrifice is by decapitation without any form of anesthesia, to minimize the potential for release of testosterone during CO_2 asphyxiation or interference with hormone levels if performed correctly. Decapitation is also considered more humane than CO_2 asphyxiation. If CO_2 is used, the Guideline recommends it must be given for no more than 60 seconds prior to decapitation, even if the animal has not fully succumbed in that time. Decapitation has generally not been found to interfere with the integrity of the thyroid, which must be maintained in order to obtain thyroid weight and histology sections. The use of anesthetic (injectable or inhalational) and exsanguination via other methods would not necessarily be a basis on which to reject a study, subject to the following caveats:

(1) Dose levels of anesthetic must be such that the majority of animals reach deep anesthesia within 2 minutes. For animals not reaching deep anesthesia within 2 minutes, either decapitate immediately or record the time until deep anesthesia is achieved and mark the animal as a deviation. Examine whether the additional time resulted in differences in hormone levels, and either use or exclude the information, as appropriate, for further analyses.

(2) The amount of blood collected via the method chosen is sufficient to perform the necessary hormonal and blood chemistry work. Use of a method that frequently does not yield sufficient blood for the necessary analyses is not likely to be acceptable.

The main concern with use of anesthetic or asphyxiant is the induction of stress, which may affect hormone levels within a short period of time. Use of injectable anesthetic is preferred due to better delivery control and thus potentially shorter times to induce deep anesthesia than typically occurs with inhalational anesthetics.

The Guideline recommends the order of necropsy be randomized or otherwise evenly distributed across all groups being necropsied that day. When two or more test chemicals use the same control group, it is particularly important to intersperse the control animal necropsies across the entire time span in which all of the necropsies for all the test chemicals and dose levels are conducted.

It is recommended that blood from the trunk of the animal be collected immediately into serum separation tubes, centrifuged and the serum stored at -20°C or colder for subsequent hormone and blood chemistry measurements.

The Test Guideline recommends that the following procedures be used during the necropsy. Organs to be collected include: both testes, epididymides, ventral prostate, dorsolateral prostate, seminal vesicle with coagulating glands and fluid, levator ani plus bulbocavernosus (LABC) muscles, thyroid (with attached portion of trachea), liver, kidneys, pituitary, and adrenals. Fresh wet weights are generally reported for each organ, with the exception of the thyroid/trachea. The thyroid is typically weighed after fixation. The kidneys and adrenals are generally weighed as pairs, and the left and right testes and epididymides are generally weighed individually. The seminal vesicle with the coagulating glands is also weighed, typically with and without fluid present. Small tissues such as the adrenals and pituitary, as well as tissues that contain fluid, are typically weighed immediately to prevent tissues from drying out prior to weighing.

The Guideline recommends thyroid (with attached portion of the trachea) and a single testis and epididymis from each animal be evaluated histologically. Either the left or the right testis/epididymis may be chosen for a study but the choice needs to be applied consistently to all animals in the study, and the choice reported. It is recommended that the testis and epididymis be fixed in Davidson's or Bouin's solution (but no longer than 24 hours) and stored in 70% ethanol until embedded in paraffin. Bouin's solution is the preferred fixative, but testis and epididymis may also be fixed in a 10% buffered formalin solution for no more than 24 hours. Tissues are then stained with hematoxylin and eosin (H&E) for histological evaluations. The thyroid, with attached trachea, is fixed in 10% buffered formalin for at least 24 hours and the

thyroid (with parathyroids) is dissected away from the trachea, weighed, and stored in 70% ethanol until embedding and staining. The kidney is also fixed in 10% buffered formalin for at least 24 hours and stored in 70% ethanol until embedding and staining.

2. Hormone Assays

Hormonal measurements can be conducted using radioimmunoassay (RIA), immunoradiometric assay (IRMA), enzyme-linked immunosorbent assay (ELISA), or time-resolved immunofluorescent procedures. The Guideline recommends that regardless of which is used, multiple quality control (QC) samples run in duplicates be included among the test samples. The Guideline recommends that any measurement kit that is used be shown to yield appropriate values for control rats at the laboratory performing the pubertal assay. This includes demonstrating that QC was performed as described by the kit manufacturer and that the performance falls within the range defined by the manufacturer. If the kit does not provide or specify a standard control, then an additional option would be for the lab to use its own historical quality control samples. The DER should generally include the performing lab's criteria for evaluating the kit's performance. If the laboratory has never had experience with the kit for making measurements specifically in the rat, EPA recommends that it test the kit with one or more untreated rats outside of the pubertal assay before relying on it for the full study, and that the findings are documented in the study report.

3. Blood Chemistry

Any standard panel of blood chemistry tests that includes creatinine and blood urea nitrogen would be scientifically appropriate as long as the measurements are calibrated for rats and the normal ranges for controls are reported. The normal ranges for controls may be from the literature (in which case the reference should be given in the DER), or from historical controls. Clinical chemistry parameters are typically performed on all animals, including controls, and are measured at terminal sacrifice.

Clinical chemistry levels are usually considered adverse when at least two liver parameters have a dose dependent, biologically significant change in albumin, alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, bilirubin, cholesterol or gamma glutamyltransferase. Typically, these changes should corroborate each other and be consistent with the known significance of the parameters. With renal toxicity, serum creatinine concentrations tend to parallel changes in blood urea nitrogen (BUN). Thus, in well-controlled toxicity studies in rodents, relatively small increases in serum BUN and creatinine concentrations (e.g., 1.5-fold) can be indicative of renal injury; but significant and consistent increases in BUN or creatinine above control ranges, including laboratory reference ranges, provide more support for a treatment related effect.

4. Histology

The Guideline recommends testis, epididymis, thyroid, and one kidney be evaluated for pathologic abnormalities and potential treatment-related effects. Thyroid sections are subjectively evaluated for follicular cell height and colloid area using a five point grading scale (1 = shortest/smallest; 5 = tallest/largest) and any abnormalities/lesions are noted. The Guideline recommends that a minimum of two sections of each of the two lobes of the thyroid be

evaluated. The Guideline provides example photomicrographs illustrating the magnitude of differences that are typically evaluated as separate scores. Guidance related to the histological evaluation of the testis and epididymis is given in EPA's Health Effects Test Guideline OPPTS 870.3800: Reproduction and Fertility Effects (US EPA, 1998): "Besides gross lesions such as atrophy or tumors, testicular histopathological examination should be conducted in order to identify treatment-related effects such as retained spermatids, missing germ cell layers or types, multinucleated giant cells, or sloughing of spermatogenic cells into the lumen .... Examination of the intact epididymis should include the caput, corpus, and cauda, which can be accomplished by evaluation of a longitudinal section, and should be conducted in order to identify such lesions as sperm granulomas, leukocytic infiltration (inflammation), aberrant cell types within the lumen, or the absence of clear cells in the cauda epididymal epithelium."

H. Statistical Analyses

The reviewer should consider whether there are any data points that should be excluded from the data set, and whether any data points that are identified as statistical outliers should actually not be excluded. The decision to exclude data points should be based on statistical analysis and toxicological judgment. Values due to obvious technical errors should be excluded. The Guideline recommends the study report justify and report outliers in the raw data; the DER should explain the exclusion of any data points.

The Test Guideline recommends that the following procedures be used for statistical analyses. For additional guidance, please refer to the Corrections and Clarifications on Technical Aspects of the EDSP Tier 1 Assays (OCSPP Test Guideline Series 890) document. This document may be found by way of the EDSP web page (http://www.epa.gov/endo/). It is recommended that the data be analyzed for normal distribution and heterogeneity of variance in order to satisfy the assumptions of Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA). The Guideline recommends that all data except histology evaluations (i.e., initial body weight [PND 23], age and body weight at preputial separation, body weight, body weight gain, and organ weights at necropsy, and serum hormones) be analyzed by ANOVA.

For calculation of body weight gain, the Guideline recommends use of the body weight on the last day all the animals were weighed; specifically, if sacrifices were performed over two days, it would not be appropriate to use the day when only the last half of the animals were available. If the study was conducted in blocks, then the Guideline recommends that the analysis be performed using a two-way ANOVA with Block and Treatment as main effects.

Age and body weight at preputial separation and all organ weights are also typically analyzed by ANCOVA, using the body weight at PND 21 as the covariate. When statistically significant effects are observed ($p < 0.05$), it is recommended that treatment means be examined further using the appropriate pairwise comparison tests that would be needed in order to identify dose groups that are significantly different from the control group. Where there is heterogeneity of variance, the recommended procedure is to transform data appropriately prior to ANOVA/ANCOVA, or to analyze the data using an appropriate nonparametric test. However, non-parametric analysis is the method of last resort as it does not allow analysis of covariance. In addition to ANOVA and ANCOVA, it is recommended that the unadjusted and adjusted values are typically examined for linear trend with dose level.

In cases where PPS has not occurred prior to necropsy for an individual animal, the last day of observation +1 is generally used as the age of PPS for purposes of calculating the mean PPS for each group. For example, if the animal was sacrificed on PND 53 without preputial separation, PND 54 is used as the value for that animal when determining the mean for the treatment group.

IV. STUDY INTERPRETATION

The following sections and text summarize the information generally expected to be obtained by the study that would generally be relevant to evaluating the pubertal assay and therefore need to be included in each DER. The Guideline recommends the study report provide all the raw data and data summaries in electronic format (spreadsheet or comma-separated values), along with all formatting information that is necessary to read the data. In circumstances where the reviewer independently calculates summaries and statistics for endpoints, it is recommended that these electronic summaries be used to reduce errors. Use of the electronic raw data for these purposes, if provided/used, should be documented in the DER. The DER should also have an executive summary describing the number and strain of rats used in the study, the dose levels and chemicals tested, and the effects, with levels of statistical significance for all endpoints.

A. Results

1. General Growth and Preputial Separation

The DER should generally include the mean body weight ± standard deviation (SD) for each day during dosing for each treatment group, including vehicle control. The DER should generally provide the data in tables. In general, the p-value is reported in the DER when the difference is statistically significant. The numeric data for general growth and PPS should typically also be provided in the DER in tabular form and include the mean, SD, coefficient of variation (CV), number of animals (N), and p-value for the endpoints listed below, for each test-chemical dose group and control, both unadjusted (U) and adjusted (A) for body weight on PND 23.

- Age at attainment of PPS
- Body weight at the age of attainment of PPS
- Initial body weight (PND 23)
- Body weight on the last day all the animals were weighed
- Final body weight as percent of control
- Body weight gain from first dose to the (first) day of necropsy

Endpoints that show an effect (by ANOVA/ANCOVA or a non-parametric test) should be indicated in the DER. The DER should list any transformations used to eliminate heterogeneity of variance, or state if the non-parametric test was used. The pairwise test used to compare the means of dosed groups to the mean of controls should also be indicated.

The incidence of the proportion of animals in which PPS did not occur by necropsy (e.g., X/15) should be reported in the DER. For these animals, a value of necropsy day +1 (e.g., PND 54) should be used for calculating the group mean.

The summary table that should be used to report general growth and PPS data in the DER is shown below in Table 1. Table 2 provides the proportion of animals that have not undergone preputial separation. Dose group parameters that are significantly different from the vehicle control group ($p < 0.05$) should be indicated.

Table 1. General Growth and Preputial Separation (PPS)[a].

Parameter Evaluated		Vehicle Control				Low Dose (# mg/kg/day)				High Dose (# mg/kg/day)			
		Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV
Initial body weight (PND 23; g)	U												
	A												
Body weight at PPS (g)	U												
	A												
Final body weight (g)	U												
	A												
Final body weight (% of control)	U												
	A												
Body weight gain (final – initial; g)	U												
	A												
Age at PPS (PND)	U												
	A												

a Data were obtained from page [#] of the study report, and are the average of [#] male rats per dose group.
U = Unadjusted for body weight on PND 23
A = Adjusted for body weight on PND 23
SD = Standard Deviation
CV = Coefficient of Variation
* Significantly different from controls at $p<0.05$.

Table 2. Proportion Unseparated

	Vehicle Control	Low Dose ([#] mg/kg/day)	High Dose ([#] mg/kg/day)
Number of Animals Examined			
Incidence			

2. Organ Weights at Necropsy

The DER should report the mean, SD, CV, number of animals (N), and p-value (for significant changes only) for both unadjusted (U) and adjusted (A; for body weight on PND 21) organ weights for the following organs: liver, kidneys, pituitary, adrenals, seminal vesicle plus coagulating glands (with and without fluid), ventral prostate, dorsolateral prostate, LABC, epididymides (left and right), testes (left and right), and thyroid. The mean, SD, and p-value (for significant changes only) of the organ-weight-to-body-weight ratio for liver, kidney, adrenals, and pituitary should also be provided in the DER. For all other organ weights, do not use relative organ to body weight ratios, and do not adjust for body weight at necropsy.

The summary table that should be used to report organ weight data in the DER is shown below in Table 3. Dose group parameters that are significantly different from the vehicle control group ($p < 0.05$) should be indicated.

Table 3. Organ Weights at Necropsy[a].

Organ		Vehicle Control				Low Dose ([#] mg/kg/day)				High Dose ([#] mg/kg/day)			
		Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV
Liver (g)	U												
	A												
	R												
Kidneys (g)	U												
	A												
	R												
Pituitary (mg)	U												
	A												
	R												
Adrenals (mg)	U												
	A												
	R												
Seminal vesicle + coagulating gland, with fluid (mg)	U												
	A												
Seminal vesicle + coagulating gland, without fluid (mg)	U												
	A												
Ventral prostate (mg)	U												
	A												
Dorsolateral prostate (mg)	U												
	A												
LABC (mg)	U												
	A												
Epididymis, left (mg)	U												
	A												
Epididymis, right (mg)	U												
	A												
Testis, left (mg)	U												
	A												
Testis, right (mg)	U												
	A												
Thyroid, fixed (mg)	U												
	A												

a Data were obtained from page [#] of the study report, and are the average of [#] male rats per dose group.
U = Unadjusted for body weight on PND 23
A = Adjusted for body weight on PND 23
SD = Standard Deviation
CV = Coefficient of Variation
R = Organ-to-body weight ratio (relative to body weight)
* Significantly different from controls at $p<0.05$.

3. Thyroxine, Thyroid Stimulating Hormone, and Testosterone Levels, and Clinical Chemistry

The mean, SD, CV, number of animals, and p-value (for significant changes only) for the thyroxine (T_4) thyroid stimulating hormone (TSH), and testosterone levels, for each treatment group, including vehicle control should be reported in the DER. Similarly, the mean, SD, CV, number of animals, and p-value (for significant changes only) for each of the clinical chemistry parameters measured should be presented. Also, the normal range for each parameter should be provided in the study report and appropriate data appended to the DER. A comparison should be made by the Reviewer for the normal values to those observed during the study.

The summary table that should be used to report hormone levels and clinical chemistry data in the DER is shown below in Table 4. Dose group parameters that are significantly different from the vehicle control group ($p < 0.05$) should be indicated.

Table 4. Hormone Levels and Clinical Chemistry [a]

Parameter Evaluated	Vehicle Control				Low Dose (# mg/kg/day)				High Dose (# mg/kg/day)			
	Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV	Number of Animals Examined	Mean	SD	CV
Hormones												
Serum T_4, Total (µg/dL)												
Serum TSH (ng/mL)												
Serum testosterone (ng/mL)												
Clinical Chemistry												
Creatinine (µmol/L)												
Blood urea nitrogen (mmol/L)												

a Data were obtained from page [#] of the study report, and are the average of [#] male rats per dose group.
SD = Standard Deviation
CV = Coefficient of Variation
* Significantly different from controls at $p<0.05$.

4. Histopathology

The testis, epididymis, thyroid, and one kidney are evaluated for pathologic abnormalities and potential treatment-related effects. Thyroid sections should be subjectively evaluated for follicular cell height and colloid area, preferably using a five point grading scale (1 = shortest/smallest; 5 = tallest/largest) and any abnormalities/lesions should be noted. The Guideline recommends a minimum of two sections of each of the two lobes of the thyroid be examined in order to obtain representative sample of the thyroid tissue from each lobe.

The judgment of histopathological changes observed on the testis, epididymides, and/or thyroid glands are important to consider when evaluating the organ weights and hormone levels measured in this assay. Severity and incidences of effects and dose-response relationship may also be important information to consider. A summary of the incidence data for histological

findings should be provided in the DER (as in Table 5), whereas the study report may additionally provide example photomicrographs of significant observations for the record.

Table 5. Incidence of Histopathological Lesions of the Thyroid Gland [a]

Treatment Groups	Parameter Evaluated											
	Colloid Quality			Follicular Cell Height (Increase)			Follicular Cell Height (Decrease)			Follicular Cell Shape		
	Severity[b]	Incidence		Severity	Incidence		Severity	Incidence		Severity	Incidence	
		O	E		O	E		O	E		O	E
Vehicle Control	0			0			0			0		
	1			1			1			1		
	2			2			2			2		
	3			3			3			3		
Low Dose (# mg/kg/day)	0			0			0			0		
	1			1			1			1		
	2			2			2			2		
	3			3			3			3		
High Dose (# mg/kg/day)	0			0			0			0		
	1			1			1			1		
	2			2			2			2		
	3			3			3			3		

a Data were obtained from page [#] of the study report.
b Thyroid histopathology is graded 0 – 3 based on severity: 0=Not remarkable, 1=Mild, 2=Moderate, 3=Severe. See OECD No. 82 for reference.
O = Number Observed
E= Number Examined

Table 6. Incidence of Histopathological Lesions of the Testes, Epididymides and Kidney [a]

Findings	Dose Level (# mg/kg bw/day)					
	Vehicle Control		Low (#)		High (#)	
	# Observed	# Examined	# Observed	# Examined	# Observed	# Examined
Testes						
[observation type]						
Epididymides						
[observation type]						
Kidney						
[observation type]						

a Data were obtained from page [#] of the study report.

B. Data Interpretation Procedure

The Guideline recommends that the highest dose level be at or just below the Maximum Tolerated Dose (MTD) but need not exceed the Limit Dose (1000 mg/kg/day); however, typically, the Agency also considers the toxicity profile of the chemical (i.e., cholinesterase inhibition, target organ toxicity, etc) in dose selection. Typically the second dose level is spaced to produce a lesser degree of toxicity relative to the high dose unless justification is provided for testing at a different level.

Generally, negative results for interaction with the endocrine system in the pubertal assay will require demonstration that the highest dose level tested was at or near a dose that was previously determined to produce toxicity. To determine the acceptability of a study in which a positive finding occurs only at a dose level that causes clear adverse effects (e.g. $\geq 10\%$ decrease in body weight gain at termination compared to controls) a weight-of-evidence assessment may be required.

As with the standard practice for evaluating changes in organ weights for the pituitary, liver and kidney, EPA recommends that terminal body weights be used to calculate relative organ weight changes. In contrast, because endocrine-active agents themselves may have an effect on body weight, it is typically appropriate to adjust for covariance with body weight before chemical treatment began (e.g. PND 23) for the accessory sex organs.

Changes in hormones and histopathology may occur independent of changes in organ weights. The severity and incidence of effect(s) are important considerations.

As with other standard evaluation procedures for OCSPP Test Guideline studies, the concurrent control is generally the first order comparison for the treatment groups. The laboratory may also choose to provide appropriate historical control data based on equivalent study parameters (strain, age, endpoints). The adequacy of the concurrent control data may be evaluated by comparison to the historical control data and/or the performance criteria provided in the Test Guideline. The data for the concurrent controls should generally fall within the range of specified performance criteria (see Table 7 below).The performance criteria are indications of whether the sensitivities of individual endpoints are sufficient to allow conclusions that the test chemical did not affect those endpoints.

The data from the male pubertal assay provide general profiles of changes in endpoints for various hormone pathways such as androgen agonism or antagonism, alteration of steroidogenesis, thyroid toxicity, and interference with HPG function. These profiles can be used to establish a “weight of evidence” for general interaction of a test chemical with the endocrine system. For example, an anti-androgen such as vinclozolin delays puberty, impairs reproductive tract development (e.g., decreased VP, SV, LABC, epididymis weight), and increases testosterone at higher doses, so a test chemical with similar responses would likely be suspected of having an anti-androgenic interaction. A similar profile would be expected if the compound inhibits testosterone synthesis. One way to discern a compound that inhibits steroidogenesis from one that interferes with androgen receptor binding is to evaluate serum testosterone (a required endpoint) as this endpoint will obviously be decreased.

Additional guidance on data interpretation are provided in the OCSPP Guideline 890.1500 for the Male Pubertal Assay and in the Agency's document titled "Corrections and Clarifications on Technical Aspects of the EDSP Tier 1 Assays (OCSPP Test Guideline Series 890)" which can be found on the EDSP web page (http://www.epa.gov/endo/).

C. Performance Criteria

The following performance criteria have been established for the vehicle-control animals. The objective of the performance criteria is to ensure the sensitivity of the endpoints relative to other confounding factors during the course of the study (e.g. phytoestrogens in diet, contaminants in drinking water). See the Data Interpretation Procedure above for correct use of the performance criteria. Units for the endpoints are shown in Table 7. The "Mean," "2 SDs," "CV," and "1.5 CV" columns describe the mean, two standard deviations, coefficient of variation, and 1.5 times the coefficient of variation for the given endpoints in the historical controls. The targeted mean values and CVs for the vehicle control group for various endpoints are provided below.

Table 7 below is intended to be used as a worksheet for comparison of the observed results to the performance criteria. Report the relevant findings in the DER.

Table 7. Performance Criteria for Controls (Male Sprague-Dawley and Wistar Strains)

Endpoint	Strain	Mean	2 SDs	Acceptable Range	CV	1.5 CV	Top of Acceptable Range[a]
Ventral prostate (g)							
	Wistar	0.223	0.072	0.151 to 0.295	16.67	5.65	22.32
	Sprague-Dawley	0.246	0.086	0.160 to 0.332			
LABC[b] (g)							
	Wistar	N/A	N/A	N/A	15.77	11.33	27.10
	Sprague-Dawley	0.651	0.204	0.447 to 0.855			
Epididymis (g)							
	Wistar	0.474	0.124	0.350 to 0.598	10.94	5.45	16.39
	Sprague-Dawley	0.446	0.082	0.364 to 0.528			
Seminal vesicle (g)							
	Wistar	0.576	0.234	0.342 to 0.810	20.61	0.45	21.06
	Sprague-Dawley	0.507	0.212	0.295 to 0.719			
Testis (g)							
	Wistar	1.341	0.250	1.091 to 1.591	9.27	8.35	17.62
	Sprague-Dawley	N/A	N/A	N/A			
T_4 (total; µg/dL)							
	Wistar	5.478	2.164	3.314 to 7.642	18.27	9.20	27.46
	Sprague-Dawley	5.716	1.660	4.056 to 7.376			
Thyroid weight (mg)							
	Wistar	N/A	N/A	N/A	15.39	8.24	23.63
	Sprague-Dawley	20	6	14 to 26			
TSH (ng/mL)							
	Wistar	N/A	N/A	N/A	34.04	24.26	58.29
	Sprague-Dawley	14.162	9.950	4.212 to 24.112			
Age at preputial separation (PND, where day of birth = PND 0)							
	Wistar	43.124	2.948	40.176 to 46.072	3.64	2.03	5.67
	Sprague-Dawley	43.147	3.366	39.781 to 46.513			
Weight at preputial separation (g)							
	Wistar	209.142	31.850	177.292 to 240.992	7.54	0.03	7.57
	Sprague-Dawley	222.223	33.946	188.277 to 256.169			
Testosterone (ng/mL)							
	Wistar	2.118	2.540	0 to 4.658	58.82	30.88	89.70
	Sprague-Dawley	2.110	1.850	0.260 to 3.960			
Final body weight (g)							
	Wistar	291.818	41.578	250.24 to 333.396	6.62	0.85	7.47
	Sprague-Dawley	295.647	36.412	259.235 to 332.059			
Adrenals (mg)							
	Wistar	54.597	13.768	40.829 to 68.365	15.42	7.34	22.77
	Sprague-Dawley	46.487	14.636	31.842 to 61.114			
Kidneys (g)							
	Wistar	2.516	0.550	1.966 to 3.066	9.56	5.20	14.76
	Sprague-Dawley	2.646	0.404	2.242 to 3.050			
Liver (g)							
	Wistar	14.070	2.874	11.196 to 16.944	10.24	4.69	14.93
	Sprague-Dawley	12.670	2.680	9.990 to 15.350			
Pituitary (mg)							
	Wistar	8.051	1.934	6.117 to 9.985	12.14	3.83	15.98
	Sprague-Dawley	10.354	2.544	7.810 to 12.898			
Weaning weight (g)							
	Wistar	58.238	11.058	47.180 to 69.296	8.04	2.21	10.25
	Sprague-Dawley	52.642	7.170	45.472 to 59.812			

a Bottom of the acceptable range for coefficient of variation is zero.
b LABC = levator ani/bulbocavernosus muscle complex.
SD = Standard Deviation; CV = Coefficient of Variation
N/A = Not available

V. CHARACTERIZATION OF FINDINGS

On completion of the review of this assay, the Agency will conduct a weight of evidence analysis to consider the potential of the chemical to disrupt the estrogen, androgen, or thyroid hormone systems. Chemicals with demonstrated evidence of a potential to interact with the estrogen, androgen, and/or thyroid hormone systems will be considered as candidates for Tier 2 testing.

VI. DATA EVALUATION RECORD

Once the study has been reviewed using the principles described in the previous sections of this SEP, a DER will be prepared. A DER template that provides additional guidance for the preparation of the DER is available.

VII. REFERENCES

Corrections and Clarifications on Technical Aspects of the EDSP Tier 1 Assays (OCSPP Test Guideline Series 890) in the docket; the link to this document may be found by way of the EDSP web page (http://www.epa.gov/endo/)

Markaverich BM, *et al*. (2002) Identification of an endocrine disrupting agent from corn with mitogenic activity. *Biochem. Biophys. Res. Commun.* 291(3): 692-700.

OECD (2007) Guidance Document on Amphibian Thyroid Histology. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 82. Paris.

USEPA, Office of Prevention, Pesticides, and Toxic Substances. August 1998. Health Effect Test Guidelines 870.3800: Reproduction and fertility effects. EPA 712-C-98-208.