Test Method for Assessment of Pubertal Development and Thyroid Function in Juvenile Female Rats, for Potential Use in the Endocrine Disruptor Screening Program

I. Purpose and Applicability

The purpose of this protocol is to quantify the effects of chemicals on pubertal development and thyroid function in the intact juvenile/peripubertal female rat. This assay detects chemicals that display anti-thyroid or estrogenic/anti-estrogenic activity (e.g., alterations in receptor binding or steroidogenesis), or alter hypothalamic function or gonadotropin or prolactin secretion.

II. Endpoints

- Growth (daily body weight)
- Age and body weight at vaginal opening
- Organ weights:
  - Uterus (blotted)
  - Ovaries (paired)
  - Thyroid
  - Liver
  - Kidneys (paired)
  - Pituitary
  - Adrenals (paired)
- Histology
  - Uterus
  - Ovary
  - Thyroid (colloid area and follicular cell height)
  - Kidney
- Hormones
  - Serum thyroxine (T_4), total
  - Serum thyroid stimulating hormone (TSH)
- Estrus cyclicity
  - Age at first estrus after vaginal opening
  - Length of cycle
  - Percent of animals cycling
  - Percent of animals cycling regularly

III. General Conditions

Rats are housed in clear plastic cages (approximately 20 x 25 x 47 cm) with heat-treated[1] laboratory-grade pine shavings as bedding. Corn cob bedding should not be used due to its potential to disrupt endocrine activity[2]. Wire mesh-bottomed caging should not be used due to the potential for pup loss.

Prior to the onset of the study, pregnant female rats are housed individually. At weaning, pups are housed in groups of either two or three females of the same treatment group per cage.

Animals are maintained on a balanced laboratory diet[3] and water[4] *ad libitum*, in a room with a 14:10 hour light:dark photoperiod (on at 0500 hours, off at 1900 hours local time), temperature of 20 to 24° C, and a relative humidity of 40 to 50%. Small variations of this portion of the protocol except for the light cycle are permissible but must be documented[5].

IV. Animals: Juvenile Female Rats

The Sprague-Dawley strain of rats is the preferred strain for this assay until a more-appropriate strain (or set of strains) is identified and associated performance criteria 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.

Juvenile female rats are derived from individually housed pregnant females that were bred in-house or purchased from a supplier as "timed pregnant" dams. All dams must be pregnant for the first time and timed to deliver on the same day. If purchased from a supplier, all dams should be on the same gestation day (GD) but that GD may be GD 7, 8, 9, or 10 at the time of arrival at the performing laboratory (where GD 0 = day of sperm positive). Dams are allowed to deliver their pups naturally. Any litters with fewer

---

[1] to eliminate resins that induce liver enzymes.

[2] Markaverich BM, Alejandro MA, Markaverich D, et al. 2002. Biochem Biophys Res Commun 291(3):692-700.

[3] N.B.: Totally synthetic diets are not appropriate, genistein-equivalent content of genistein plus daidzein (aglycone forms) of each batch must be less than approximately 300 ug/g, and the same batch of feed must be used for treated and control groups at all times. ("Genistein-equivalent content" of daidzein is approximately 0.8. Owens WB, Ashby J, Odum J, Onyon L. 2003. Environ Health Perspect 111(12):1559-1567.)

[4] Deionized water is preferred. Tap water is not acceptable.

[5] EPA recognizes that the temperature and humidity ranges specified are significantly narrower than specified in guidelines published by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). However, EPA regards the AAALAC guidelines as exceptionally wide and the extremes of that range only marginally tolerable for the pubertal assay. Efforts should be made to keep the conditions within the range specified above.

than 8 total pups/litter (i.e., including both males and females) and any litters not delivered by GD 23 are excluded from the study. To maximize uniformity in growth rates, the litters are standardized to 8 to 10 pups per litter between post-natal days (PND) 3 and 5. (PND 0 is defined as the day on which the pup is first seen, assuming that the cages are checked for new births daily, in the morning.) Reducing litter size to 8-to-10 is required when dams have more than enough pups, but cross-fostering to raise litter size to the required number is not acceptable. Body weights are monitored weekly and any unthrifty litters or runted pups excluded from the study. Enough litters should be available to assure that a sufficient number of juvenile females are available for 15 female pups per treatment group. (If parallel male and female pubertal studies are being conducted, males and females from the same litter can be used in their respective studies.)

The pups are weaned on PND 21. Also on day 21, female pups are marked by litter, and then all of the female pups from all of the litters are weighed individually to the nearest 0.1 g and ranked by body weight. A population of female pups that is as homogeneous as possible is selected for the study by eliminating an equal number of pups from the heavy end and the light end of the distribution, leaving the number of animals needed for the study in the middle. In this way, one nuisance variable (viz., body weight at weaning) is experimentally controlled. The female pups are assigned to treatment groups such that the mean body weights and variances for all groups are similar. Avoid placing littermates in the same group.

After assignment to treatment groups, female offspring belonging to the same treatment group are housed in groups of either 2 or 3 per cage, such that each cage has the same number of animals. (In the case of housing 2 per cage and a planned N of 15, it will be necessary to add an additional rat to the last cage.)

It is imperative that treatments be initiated no later than PND 22. Waiting just a few days longer can result in failure of the study as the onset of puberty (i.e., vaginal opening) in the control female rats will begin within a few days.

As described in more detail later in this document, the preferred procedure is to kill all the animals on a single day to close the in-life portion of the study, but kills may be conducted over two days rather than one, with half of each group killed on each day. If kills over two days are planned, the kill days are assigned at the time of initial distribution to treatment groups and the body weights are distributed across kill groups such that the mean body weights and variances for all groups are similar, and, when possible, litter mates are not in the same group.

V. Experimental Design

This protocol uses a randomized complete block design (time-separated necropsy is the blocking factor) with fifteen female rats in each treatment group. The treatment

conditions are (1) vehicle-treated and (2) xenobiotic-treated (two dose levels). The highest dose level should be at or just below the Maximum Tolerated Dose (MTD) level but need not exceed the limit dose of 1 g/kg/day. A dose level will be considered to be at or just below the Maximum Tolerated Dose level if it causes a statistically significant reduction in terminal body weight in treated animals vs. controls, the reduction is no greater than approximately 10% of the mean for the controls, and no clinical signs of toxicity associated with the dose level are observed throughout the study. In addition, abnormal blood chemistry values at termination (particularly creatinine and blood urea nitrogen (BUN)) may indicate that MTD was exceeded, even in the absence of a reduction in terminal body weight compared to controls. Histopathology of the kidney (or any other organ where gross observations indicate damage) may be used as evidence that MTD was exceeded. Blood chemistry and histopathology of the kidney are not required, however.

The second dose level should be one half of the highest dose level being tested unless justification is provided for testing at a different level.

If necessary, the study can be conducted in time-separated blocks rather than at one time. In this case, each block should contain all treatment groups and be balanced with respect to numbers of animals and body weight at weaning.

VI. Test substance

Chemical purity and stability in vehicle must be known prior to testing so that dose levels are correctly prepared.

Corn oil is the preferred vehicle for all treatment groups. If corn oil is used, it must be clear and free of sediment. It should have a bland odor, free from rancid, musty, metallic, putrid or any other undesirable odor. Other solvents such as water or carboxymethylcellulose may be used where appropriate. If the test substance is not soluble in any of the conventional solvents, it is administered as a suspension in corn oil. 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..

VII. Treatment

Each animal is weighed daily, prior to treatment, and the body weight recorded. Clinical observations are also recorded daily. Animals which are found dead or which must be euthanized *in extremis* are removed from the cage. Endpoint measures (organ weights, hormone levels, histology, etc.) are not taken from these animals.

Treatments are administered daily by oral gavage from PND 22 through PND 42. This duration of treatment is unnecessary to detect estrogenic chemicals, but is required for the detection of pubertal delay and antithyroid effects. Test chemicals are administered in 2.5 to 5.0[6] ml corn oil/kg body weight at 0700-0900 daily using an 18 gauge gavage needle (1 to 1½ inch length, 2.25 mm ball) and a 1 cc (disposable) tuberculin syringe for each treatment. Needle gauge may be optimized to animal size but must be constructed of metal to avoid the potential for absorption by or leaching of substances from rubber or plastic tubing. The treatments are administered on a mg/kg body weight basis using the current day's weight, and volume of the dose administered is 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 not 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. As general guidance, EPA suggests that a reduction in body weight when compared to controls of more than 20% in the absence of other signs of toxicity *may* justify removal.

VIII. Vaginal Opening

Beginning on PND 22, females are examined daily for vaginal opening. The appearance of a small “pin hole”, a vaginal thread, and complete vaginal opening are all recorded on the days they are observed. The day of complete vaginal opening is the endpoint used in the analysis for the age at vaginal opening. However, if any animal within any treatment group shows incomplete opening (such as persistent threads or a "pin hole") for greater than three days, a separate analysis is conducted using the ages at which incomplete opening was first observed. Documentation of a vaginal thread even if vaginal opening otherwise appears complete is important. It is also critical that “initiation” of vaginal opening be recorded. It is preferred but not critical that vaginal opening observations be taken after the daily dosing. Whether collected before or after dosing, the vaginal opening observations must be collected at approximately the same time each day.

IX. Estrous Cyclicity

Beginning on the day of vaginal opening, through to and including the day of necropsy, daily vaginal smears are obtained and evaluated under a low-power light microscope for the presence of leukocytes, nucleated epithelial cells, or cornified epithelial cells. The vaginal smears are classified as diestrus (predominance of leukocytes mixed with some

[6] Dosage volume per kg body weight must be the same for all treated animals in the experiment but the value chosen for the study may be anywhere in the specified range.

cornified epithelial cells), proestrus (predominance of clumps of round, nucleated epithelial cells), or estrus (predominance of cornified epithelial cells) and the stage is recorded daily. (Metestrus is classified as (an early part of) diestrus rather than (a late part of) estrus.) Age at first estrus is noted[7]. It is preferred but not critical that estrous cycle observations be taken after the daily dosing. Whether collected before or after dosing, estrous cycle observations must be collected at approximately the same time each day.

At the end of the study, the overall pattern of each female is characterized as regularly cycling (having recurring 4- to 5-day cycles), irregularly cycling (having cycles with a period of diestrus longer than 3 days or a period of cornification longer than 2 days), or not cycling (having prolonged periods of either vaginal cornification or leukocytic smears).

X. Necropsy

Females are killed on PND 42. If necessary, one half of the females may be killed on PND 42 and the remaining females on PND 43 as long as the animals in each treatment group are equally dispersed between the two necropsy days. Animals killed on PND 43 are dosed and treated on the day of kill just like the animals killed on PND 42 with regard to time of dosing, collection of vaginal opening and estrous cycle observations, etc. All animals are dosed between 0700 and 0900 hours local time, and killed beginning 2 hours following dosing. It is critical that kills are completed by 1300 hours due to normal diurnal fluctuation in thyroid hormone levels. On the day of kills, moving the cages or otherwise stressing the animals unnecessarily should be avoided so that variations in stress-related hormone levels are minimized.

The preferred method of kill is by decapitation without any form of anesthesia, to minimize the potential for release of hormones during CO_2 asphyxiation or interference by other anesthetics. Decapitation is considered more humane than CO_2 asphyxiation. If CO_2 is used, 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 order of necropsy is randomized or otherwise evenly distributed across *all* groups being necropsied that day. That is, do *not* necropsy all animals in one group before moving to the next group. When two or more test chemicals use the same control group, it is particularly important to intersperse the control animal necropsies across the

[7] See Goldman, J.M., Murr, A.S., and Cooper, R.L. (2007) The rodent estrous cycle: characterization of vaginal cytology and its utility in toxicological studies. Birth Defects Research (Part B) 80:84-97. A typical cycle consists of two or three days of diestrus, one day of proestrus, and one or two days of estrus. In the postpubertal female, this pattern develops shortly after vaginal opening and regular cycling is the norm for the young adult female.

entire time span in which all of the necropsies for all the test chemicals and dose levels are conducted.

Blood from the trunk of the animal is collected immediately (e.g., by inversion over a funnel)[8]. After collection, the blood is centrifuged. The serum is pipetted into siliconized microcentrifuge tubes[9] and stored at -20° C or colder for subsequent thyroxine (T_4) and TSH measurements.

At necropsy, the ovaries (without oviducts), uterus, thyroid (with attached portion of trachea), liver, kidneys, pituitary, and adrenals are removed and the weights of each except the thyroid/trachea recorded in milligrams to one decimal place with the exception of kidney and liver, which are recorded in grams, to two decimal places.. (Kidneys, adrenals, and ovaries are weighed as pairs.) Care must be taken to remove mesenteric fat from the uterine horns with small surgical scissors. The uterus and cervix are separated from the vagina[10]. The uterus is then placed on a paper towel, slit to allow the fluid contents to leak out, gently blotted dry and weighed. Small tissues such as the adrenals and pituitary, as well as tissues that contain fluid, should be weighed immediately to prevent tissues from drying out prior to weighing. Measures to prevent drying out may be necessary if such organs cannot be weighed immediately.

The ovaries and uterus are placed in 10% buffered formalin for at least 24 hours, after which they are rinsed and stored in 70% ethanol until embedded in paraffin. They are then stained with hematoxylin and eosin (H&E) for subsequent histological evaluations. The thyroid, with attached trachea, is fixed in 10% buffered formalin for at least 24 hours. Then the thyroid (with parathyroids) is dissected from the trachea, blotted and weighed to the nearest 0.01 mg, placed in 70% ethanol until embedded in paraffin, stained with H&E, and histologically evaluated. Kidney, like thyroid, is fixed in 10% buffered formalin for at least 24 hours, then placed in 70% ethanol until embedded in paraffin, stained with H&E, and histologically evaluated.

XI. Hormonal Assays

Hormonal measurements can be conducted using radioimmunoassay (RIA), immunoradiometric assay (IRMA), enzyme-linked immunosorbent assay (ELISA), or time-resolved immunofluorescent procedures. Regardless of which is used, always include multiple quality control (QC) samples run in duplicates that are dispersed among the test samples. Any measurement kit that is used must be shown to yield appropriate values *for control rats at the laboratory performing the pubertal assay.* The lab's criteria for evaluating the kit's performance must be included in the study report. If the

[8] Heart puncture after no more than 60 seconds of CO_2 administration is an alternative method of obtaining blood but is less preferred due to the inhumane stress of asphyxiation and the smaller volume of blood collected.

[9] If there is a greater volume of blood than will fit in one microcentrifuge tube, prepare as many separate aliquots as appropriate before freezing. Do not freeze in large aliquots, to avoid excessive freeze/thaw cycles.

[10] See attachment for guidance.

laboratory has never had experience with the kit for making measurements specifically in the rat, it might be appropriate to test the kit in one or more untreated rats outside of the pubertal assay before relying on it for the full study.

XII. Histology

Uterus, thyroid, one ovary, and one kidney are 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)[11] and any abnormalities/lesions noted. A minimum of two sections of each of the two lobes of the thyroid are evaluated. Example photomicrographs are attached. The examples illustrate the magnitude of differences that are typically evaluated as separate scores, but the reader will need to establish the range appropriate for the particular study being evaluated.

Ovarian histology following H&E staining should include an evaluation of follicular development (including presence/absence of tertiary/antral follicles, presence/absence of corpora lutea, changes in corpus luteum development, changes in number of both primary and atretic follicles) in addition to any abnormalities/lesions, such as ovarian atrophy.

Uterine histology should document cases of uterine hyper- or hypotrophy as characterized by changes in uterine horn diameter and myometrial, stromal, or endometrial gland development.

The final histological assessment should take into account the stage of the estrous cycle of the female at the time of necropsy as ovarian and uterine cellular changes are dependent upon endocrine status.

XIII. Statistical Analysis

Consideration should be given to whether there are any data points that should be excluded from the data set, and whether any data points that look as if they are outliers by an appropriate statistical test should actually not be excluded, based on toxicological judgment. Values due to obvious technical errors are excluded. Justification for exclusion of each data point must be given. Outliers must be specified in the raw data. Do not test incidence data, e.g., from histopathology evaluation, for outliers.

[11] see Capen CC, Martin SL. 1989. The effects of xenobiotics on the structure and function of thyroid follicular and C-cells. Toxicol Pathol 17(2):266-93.

All data except histology and cyclicity evaluations (i.e., initial body weight [PND 22], body weight gain[12], age and body weight at vaginal opening, body and organ weights at necropsy, and serum hormones) are analyzed by Analysis of Variance (ANOVA). If the study was conducted in blocks, then the analysis is a two-way ANOVA with Block and Treatment as main effects. Age and body weight at VO, and all organ weights should also be analyzed by Analysis of Covariance (ANCOVA) using the body weight at PND 21 as the covariate[13]. When statistically significant effects are observed ($p < 0.05$), treatment means are examined further using appropriate pairwise comparison tests to compare the control with each dose group[14]. Where there is heterogeneity of variance, data should be transformed appropriately prior to ANOVA/ANCOVA, or analyzed using an appropriate nonparametric test. Non-parametric analysis should be the method of last resort since it does not allow analysis of covariation. In addition to ANOVA and ANCOVA, examine the unadjusted and adjusted values for linear trend with dose level.

In cases where vaginal opening has not occurred prior to necropsy, use the last day of observation +1 as the age at vaginal opening when determining the mean for each group. For example, if the animal was killed on PND 42 without vaginal opening, use PND 43 as the value for that animal when determining the mean for the treatment group.

Chi-square analysis is used to determine significant differences between the cycling status (cycling vs. not cycling) of the treated groups from the control group. Similarly, chi-square analysis is used to determine significant differences between treated groups from the control group for the percent of animals cycling regularly.

Cycle length may be defined as either the number of days from one proestrus to the next proestrus, or from one diestrus to the next diestrus. Whichever definition is chosen must be applied uniformly to all groups in the study. Incomplete cycles are not counted in calculating mean cycle length. Mean cycle length for each animal is calculated first, and the mean of these means is then calculated to represent the group.

XIV. Data Summary

Provide the following figure and tables for each test chemical along with the respective control. **Be sure to use the units shown in the example tables.** Provide values to

[12] Use the body weight on the last day all the animals were weighed. Specifically, if kills were performed over two days, do not use the day when only the last half of the animals were available.

[13] The covariate is body weight on the day of weaning rather than on the day of kill because ANCOVA assumes that the covariate that is being adjusted for is not affected by the treatment, whereas in this assay endocrine-active substances may affect the overall body weight gain and thus body weight at kill. Using body weight at kill as covariate could mask a potentially endocrine effect on an organ. The Agency understands that using body weight at kill as covariate might identify which organs are more sensitive (or less sensitive) than body weight to potentially endocrine effects, but has chosen to maximize the potential of identifying organ-specific effects rather than relying on bodyweight as an indicator of potential endocrine activity.

[14] Comparison of the means of dose groups to each other is not required.

one decimal place for the organs reported in milligrams, and to two decimal places for those organs that are required to be reported in grams.

Figure 1 - Graph the mean body weight +/- Standard Deviation (SD) for each day during dosing for each treatment group, including vehicle control. (If animals were necropsied over 2 days, do not include the body weight from the last day of necropsy since only half of the animals are available.) Place an arrow at the mean age of the controls at VO. All three groups (control, dose level 1, and dose level 2) are plotted on the same graph but are distinguished from each other by point and line styles.

Table 1 - Vaginal opening; general growth
Report the mean, standard deviation, coefficient of variation, number of animals (N), and p-value for the following endpoints, for each treatment group, both unadjusted (U) and adjusted (A) for body weight on PND 21:

a) age at VO
b) body weight at the age of VO
c) initial body weight (PND 22)
d) body weight on the last day all the animals were weighed (i.e., if kills were performed over two days, do not use the day when only the last half of the animals were available)
e) final body weight as percent of control (leave control column blank)
f) body weight gain from first dose to the (first) day of necropsy.

Mark endpoints that show an effect (by ANOVA/ANCOVA or a non-parametric test, as appropriate) with an asterisk in the "Effect" column. List the transformation (if any) used to eliminate heterogeneity of variance, or the non-parametric test used, in the "Transform or nonparam" column. Name the pairwise test used to compare the means of dosed groups to the mean of controls in the "Pairwise test" column.

Mark means that are significantly different from control means ($p < 0.05$) by shading the cell (rather than by the traditional asterisk).

Show the proportion of animals in which VO had not occurred by the time of necropsy (e.g., X/15), and explain that age-at-necropsy plus one (e.g., 43) was used for those animals when calculating the mean.

Table 1

Chemical Name		Effect	Transform or nonparam	Pairwise test	Vehicle Control				(Dose Level 1)					(Dose Level 2)				
					Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Age at VO (PND)	U																	
	A																	
Body weight at VO (g)	U																	
	A																	
Initial body weight (PND 22, g)	U																	
	A																	
Final body weight (g)	U																	
	A																	
Final body weight (% of control)	U																	
	A																	
Body weight gain (final minus initial body weight) (g)	U																	
	A																	
Proportion unopened (#/N)																		

*Means different from controls at p<0.05 are marked by a shaded cell.

Table 2 - Estrous cycle status and organ weights at necropsy:
Report the number of animals in each stage of the estrous cycle at necropsy, for each treatment group. An animal is considered to be "not cycling" if she shows three or more consecutive days of estrus or five or more consecutive days of diestrus.

Report the mean, standard deviation, coefficient of variation, number of animals (N), and p-value for liver, kidneys, pituitary, adrenals, ovaries, uterus, and thyroid weights, for each treatment group, both unadjusted (U) and adjusted (A) for body weight on PND 21. Do not use relative organ to body weight ratios, and do not adjust for body weight at necropsy.

Table 3 - Estrous cyclicity
Show the mean age at first estrus for each treatment group and mark those groups which are significantly different from the vehicle control group ($p < 0.05$). Report the mean cycle length for the group. Report also the percent of each group cycling, and the percent cycling regularly, and mark those groups which are significantly different from the vehicle control group.

Table 4 - T_4 and TSH levels
Report the mean, standard deviation, coefficient of variation, number of animals, and p-value for the T_4 and TSH levels, for each treatment group.

Prepare 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 except histology. Include a summary of the histological findings.

Electronic and hard copies of spreadsheets containing the raw data from all animals for each endpoint are to be submitted to the EPA. The full report of the histological findings with photomicrographs of significant observations must also be provided.

Table 2

Chemical name					Vehicle Control				(Dose Level 1)					(Dose Level 2)				
Cycle status at kill																		
Diestrus																		
Proestrus																		
Estrus																		
Not cycling																		
Organ weights		Effect	Transform or nonparam	Pairwise test	Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Liver (g)	U																	
	A																	
Kidneys (g)	U																	
	A																	
Pituitary (mg)	U																	
	A																	
Adrenals (mg)	U																	
	A																	
Ovaries (mg)	U																	
	A																	
Uterus, wet (mg)	U																	
	A																	
Uterus, blotted (mg)	U																	
	A																	
Thyroid (mg)	U																	
	A																	

*Means different from controls at p<0.05 are marked by a shaded cell.

Table 3.

Test article	Dose (mg/kg/day)	Mean age at first estrus (PND)	Mean cycle length (days)	Cycling (%)	Regularly cycling (%)
Vehicle					
Test chemical					

Differences from controls at p<0.05 are marked by a shaded cell.

Table 4

Chemical name				Vehicle Control				(Dose Level 1)					(Dose Level 2)				
	Effect	Transform or nonparam	Pairwise test	Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Serum T_4, total (µg/dl)																	
Serum TSH, (ng/ml)																	

*Means different from controls at p<0.05 are marked by a shaded cell.

XV. Performance criteria

The following performance criteria have been established for the vehicle-control animals. See the Data Interpretation Procedure for use of the performance criteria. Units for the endpoints are shown in the table. Coefficients of variation (CVs) are in percent. 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 that endpoint in historical controls. Mean values and CVs for the vehicle control group must fall in the acceptable range of each to be considered fully acceptable.

Table 5. Performance criteria for controls (Sprague-Dawley strain)

Endpoint	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
Uterus, blotted (milligrams)						
	298.89	111.49	187.40 to 410.38	28.26	9.48	37.73
Ovaries (milligrams)						
	75.65	39.11	36.54 to 114.77	17.37	5.83	23.20
T_4 (total, ug/dl)						
	4.03	1.34	2.69 to 5.38	21.38	8.01	29.39
Thyroid weight (milligrams)						
	14.20	8.00	6.20 to 22.20	23.97	14.61	38.58
Age at VO (postnatal day, where day of birth = PND 0)						
	33.15	2.47	30.67 to 35.62	4.69	1.82	6.52
Weight at VO (grams)						
	116.58	14.86	101.71 to 131.44	9.81	4.16	13.97
Final body weight (grams)						
	154.70	49.85	104.86 to 204.55	6.69	2.25	8.93
Adrenals (milligrams)						
	43.59	5.25	38.34 to 48.84	16.02	6.95	22.97
Kidneys (grams)						
	1.57	0.63	0.95 to 2.20	8.42	2.34	10.76
Liver (grams)						
	8.05	3.73	4.32 to 11.78	9.58	3.55	13.13
Pituitary (milligrams)						
	8.97	3.11	5.86 to 12.08	15.76	11.22	26.97

No performance criteria have been established yet for TSH since there were too few studies from which reliable historical control values resulting from the same analytical method could be obtained. Such criteria may be established in the future as more data become available.

XVI. Data Interpretation Procedure

The female pubertal assay is intended to be one of a suite of *in vitro* and *in vivo* assays for determining the potential of a substance to interact with the endocrine system (Tier 1 assays). Therefore, it is important to emphasize that the data interpretation of a specific chemical will be a combination of the results from a number of these Tier-1 screening assays taken as a whole and not merely the sum of results of assays interpreted in isolation. That said, there are certain guidelines that can be given for interpreting data from a female pubertal assay.

First, the dose levels tested should be examined to see if a Maximum Tolerated Dose was used. (The highest dose level need not exceed a limit dose of 1 g/kg/day, even if MTD has not been reached.) Body weight loss (compared to controls at termination) that does not exceed approximately 10% is an indication that MTD was approached but not exceeded. Adverse clinical observations or histopathology of the kidney and/or other organs, and/or significant deviations from standard blood chemistry values may be indications that MTD was exceeded.

Negative results for interaction with the endocrine system in the pubertal assay will generally require demonstration that the highest dose level tested was at or near the MTD. Positive results in the assay generally require no such proof, but will generally require demonstration that interference due to body weight loss *per se* was not a factor in generating the results. Studies that suggest interaction with endocrine systems only at a dose level that causes more than approximately 10% body weight loss at termination compared to controls may require additional studies and/or a weight-of-evidence approach using other information in order to be interpretable.

The endpoint values for the control group should be compared to the performance criteria. Comparison should be made on the basis of the measured values, not adjusted values. Any endpoints which do not meet the performance criteria in controls will generally be disregarded for the test chemicals if they are negative but may provide useful information if they are positive.

Information that is missing due to inability to meet a performance criterion is not the same as a negative result. The more endpoints that are missing, the less likely the study will be regarded as adequate. No firm rules can be given for the minimum number of endpoints that must be available for evaluation since some of the endpoints are somewhat redundant while others are not. In general, however, missing one or two performance criteria will not be regarded as fatal to the study.

More emphasis will be placed on meeting performance criteria for the coefficients of variation than for the endpoint control means. Laboratories may submit

historical data for their own colonies to substantiate claims that tissue weights or other endpoints in the study being evaluated are in line with historical values of controls in that laboratory.

Once the usable data set has been identified through application of the performance criteria, it is evaluated to see if there is evidence of interaction of the test chemical with the endocrine system.

Due to the covariance of certain organ weights with body weight, care should be taken in interpreting pituitary, liver, and kidney weight changes. Only if a change in the organ weight relative to body weight is significant for these particular organs (i.e., not all the organs) should the weights adjusted for covariance with body weight at weaning for these particular organs be interpreted as relevant.

Endpoints other than pituitary, liver, and kidney weights should not be evaluated based on their values relative to terminal body weight, nor should an analysis of covariance with terminal body weight be used for interpretation. Since endocrine-active agents themselves may have an effect on body weight, it is most appropriate to adjust for covariance with body weight at weaning, before chemical treatment began.

Weight of ovaries and uterus must be interpreted carefully, due to the natural variability in these endpoints in cycling animals. In general, regularity of cycling should be given more weight than lack of statistical significance for the difference in weight of ovary or uterus in treated animals compared to controls. (Presence of a statistically significant difference from controls should be considered more informative than absence of such a difference.)

Because there are multiple endpoints examined in this assay, there is redundancy for the detection of potential endocrine system interaction. For example, both strong (ethynyl estradiol) and weak (methoxychlor) estrogens dramatically advanced the age of vaginal opening, altered body weight at VO, and age at first estrus. Redundancy is particularly useful when the responses from all the redundant endpoints are consistently positive since it gives greater confidence that the interaction with the endocrine system is real. However, consistency across all redundant endpoints is not required in order to infer interaction with the endocrine system. There may be valid reasons for apparently-redundant endpoints to differ in their response.

If an isolated endpoint is positive at the lower dose and no effect is seen at the higher dose, then the effect and the overall conclusions about the substance may need to be questioned. However, since the assay requires only two dose levels, the dose-response information from the female pubertal assay is sparse and informs the weight of evidence for interaction with the endocrine system but generally does not control it.

Compounds that exert effects via various mechanisms or modes of interaction with the endocrine system can be identified using the female protocol. A summary of the kinds of effects that might be seen from various different modes of action is shown in Table 6. The table is provided to help with interpretation of results, but determining a mode of action is *not* required in order to consider the assay positive for interaction with the endocrine system. Furthermore, this table is not to be interpreted as requiring that all of the endpoints shown to respond as indicated for a particular mode. Interaction with the endocrine system may be occurring without the complete profile shown.

Table 6. Potential changes indicative of different modes of action that may be observed in the female pubertal protocol

Estrogen agonist	Inhibition of Steroidogenesis	Disruption of Hypo-pit axis[1]	Thyrotoxicants
Early VO, pseudoprecocious puberty	Delayed VO	Alterations in VO	Decreased T_4
Reduced BW at VO	Delayed first estrus	Alterations in cyclicity	Alterations in TSH
Early first estrus	Persistent diestrus	Altered ovarian, uterine or pituitary weights	Changes in thyroid histology
Altered organ histology	Reduced uterine weight	Altered organ histology	Changes in thyroid weight
Possible persistent estrus	Altered organ histology		Changes in liver weight/enzyme profile
Reduced ovarian weight			
Increased Uterine weight			

[1] Changes in hypothalamic-pituitary function may advance or delay puberty, modify the ovarian cycling by inducing early cycles, alter the regularity of cycles and alter tissue weights depending on whether the chemical activates or inhibits pubertal development..

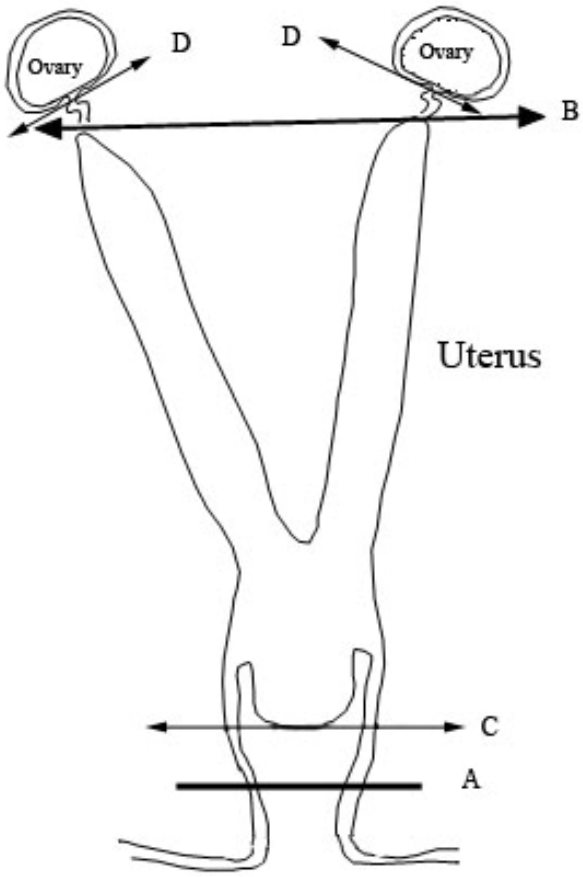

Figure 1. Removal and preparation of the uterine and ovarian tissues for weight measurement

Make a medial incision approximately five inches long on the ventral aspect of the rat from the vaginal opening towards the head. Locate the vagina ventral to the urinary bladder. Locate the uterine horns and ovaries bilaterally and detach from the dorsal abdominal wall. Detach the uterus and vagina from the body by incising the vaginal wall at A. Transfer the tissues to a tared weigh boat. Detach the ovaries by cutting between the small white oviducts and the ends of the uterine horns (B). Before measuring the uterine weight, carefully trim away the excess fat and connective tissue from the uterine horns and body and remove the vagina by cutting just caudal to the uterine cervix as shown in the figure (C). Uterine weight without luminal fluid (i.e., blotted weight) should be measured. Finally, cut between each oviduct and ovary (D), remove the ovarian bursa, and trim away the ovarian fat before measuring the weight of each ovary.