INTEGRATED SUMMARY REPORT

for

Validation of a Test Method for Assessment of Pubertal Development and Thyroid Function in Juvenile Female Rats as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery

October 2007

U.S. Environmental Protection Agency
Office of Science Coordination and Policy
and
Office of Research and Development
Washington, D.C.

Table of Contents

List of Tables

List of Appendices

Test Method for Assessment of Pubertal Development and Thyroid Function in Juvenile Female Rats for potential use in the Endocrine Disruptor Screening Program

I. Introduction

A. Purpose of the EDSP

Section 408(p) of the Federal Food Drug and Cosmetic Act (FFDCA) requires the U.S. Environmental Protection Agency (EPA) to

> *develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate [21 U.S.C. 346a(p)].*

Subsequent to passage of the Act, the EPA formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a committee of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement its Endocrine Disruptor Screening Program (EDSP). The EDSP is described in detail at the following website:

http://www.epa.gov/scipoly/oscpendo/

Upon recommendations from the EDSTAC (1998), the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormonal systems as well as wildlife.

B. Tiered approach

The EPA accepted the EDSTAC's recommendations for a two-tier screening program as proposed in a Federal Register Notice in 1998 (USEPA (1998)). The purpose of Tier 1 is to identify the potential of chemicals to interact with the estrogen, androgen, or thyroid (EAT) hormonal systems. A negative result in Tier 1 would be sufficient to put a chemical aside as having low to no potential to cause endocrine disruption, whereas a positive result would require further testing in Tier 2. The purpose of Tier 2 is to more definitively identify and characterize the potential hazard on the

endocrine system and to provide risk assessment based, in part, on dose-response relationships. Tier 2 is expected to comprise multigeneration tests in species representative of various taxa (i.e., mammals, birds, fish, amphibians, and invertebrates).

C. The Tier-1 battery

The EDSTAC (1998) concluded that a Tier-1 battery should be comprised of a suite of complementary screening assays having the following characteristics:

- Maximum sensitivity to minimize false negatives while permitting an as yet undetermined, but acceptable, level of false positives.
- Range of organisms representing known or anticipated differences in metabolic activity and include assays from representative vertebrate classes to reduce the likelihood that important pathways for metabolic activation or detoxification of parent substances or mixtures are not overlooked.
- Capacity to detect all known modes of action (MOAs) for the endocrine endpoints of concern. All chemicals known to affect the action of EAT hormones should be detected.
- Range of taxonomic groups among the test organisms. There are known differences in endogenous ligands, receptors, and response elements among taxa that may affect the endocrine activity of chemical substances or mixtures.
- Diversity among the endpoints and within and among assays to reach conclusions based on "weight-of-evidence" considerations. Decisions based on the screening battery results will require weighing the data from several assays.
- Inexpensive, quick, and easy to perform.

To detect chemicals that may affect the EAT hormonal systems through any one of the known MOAs — interruption of hormone production or metabolism, binding of the hormone with its receptor, interference with hormone transport, etc. — the EDSTAC recommended the *in vitro* and *in vivo* assays shown in Table 1 for inclusion in the Tier-1 screening battery.

Table 1. Tier-1 in vitro and in vivo screening assays recommended by the EDSTAC

Assays	Reasons for consideration
Estrogen receptor (ER) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the ER.
Androgen receptor (AR) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the AR.
In vitro steroidogenesis	A sensitive *in vitro* test to detect chemicals that interfere with the synthesis of the sex steroid hormones.
Uterotropic (rat)	An *in vivo* assay to detect estrogenic chemicals. It offers the advantage over the binding assay of incorporating absorption, distribution, metabolism, and excretion (ADME)
Hershberger (rat)	An *in vivo* assay to detect androgenic and anti-androgenic chemicals. It offers the advantage over the binding assay of incorporating ADME and differentiating between AR agonists and antagonists.
Pubertal female (rat)	An assay to detect chemicals that act on estrogen or through the hypothalamus-pituitary-gonadal (HPG) axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system.
Frog metamorphosis	A sensitive assay for detection of chemicals that interfere with the thyroid hormone system.
Fish screen	Fish are the furthest removed from mammalians among vertebrates both from the standpoint of evolution—their receptors and metabolism are different from mammals—and exposure/habitat, since they would be subject to exposure through the gills, whole body, and diet. Thus, the fish assay would augment information found in the mammalian assays and would be more relevant than the mammalian assays in triggering concerns for fish.

In addition, the EDSTAC recognized there were other combinations of screening assays that may be suitable and, therefore, recommended that the EPA validate the alternative screening assays shown in Table 2.

Table 2. Alternative in vitro and in vivo assays recommended for the Tier-1 Screening Battery

Assays	Reasons for consideration
In vitro placental aromatase	The aromatase assay detects chemicals that inhibit aromatase and would be needed if either of the two following assays using males were substituted for the female pubertal assays. The male is not believed to be as sensitive to alterations in aromatase as the female and would not therefore be sufficient to detect interference with aromatase in the screening battery.
Pubertal male (rat)	The assay detects chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.
Adult male (rat)	The assay is also designed to detect chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.

D. Validation

As noted, Section 408(p) of the FFDCA requires the EPA to use validated test systems. Validation has been defined as *"the process by which the reliability and relevance of a test method is evaluated for a particular use"* (OECD (1996); NIEHS (1997)).

Reliability is defined as the reproducibility of results from an assay within and between laboratories.

Relevance describes whether a test is meaningful and useful for a particular purpose (OECD (1996)). For Tier-1 EDSP assays, relevance can be defined as the ability of an assay to detect chemicals with the potential to interact with the EAT hormonal pathways.

Federal agencies are also instructed by the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) Authorization Act of 2000 to ensure that new and revised test methods are valid prior to their use.

In general, the EPA is following a five-part or stage validation process outlined by the ICCVAM (NIEHS (1997)). The EPA believes that it is essential to recognize that this process was specifically developed for *in vitro* assays intended to replace *in vivo* assays. The fundamental problem confronting the EPA is how to adapt and work with this process for rodent and ecological *in vivo* assays in Tiers 1 and 2 that have no suitable *in vitro* substitute.

Nonetheless, the stages of the process outlined by the ICCVAM are as follows:

First Stage - *Test Development*, an applied research function which culminates in an initial protocol. As part of this phase, the EPA prepares a Detailed Review Paper (DRP) to explain the purpose of the assay, the context in which it will be used, and the scientific basis upon which the assay's protocol, endpoints, and relevance rest. The DRP reviews the scientific literature for candidate protocols and evaluates them with respect to a number of considerations, such as whether the candidate protocols meet the assay's intended purpose, the costs and other practical considerations. The DRP also identifies the developmental status and questions related to each protocol; the information needed answer the questions; and, when possible, recommends an initial protocol for the initiation of the second stage of validation.

Second Stage - *Standardization and Optimization,* in which the protocol is refined, optimized, standardized and initially assessed for transferability and performance. Several different types of studies are conducted during this second phase depending upon the state of development of the method and the nature of the questions that the protocol raises. The initial assessment of transferability is generally a trial in a second laboratory to determine that another laboratory besides the lead laboratory can follow the protocol and execute the study.

Third Stage - *Inter-laboratory Validation* studies are conducted in independent laboratories with the optimized protocol. The results of these studies are used to determine inter-laboratory variability and to set or cross-check performance criteria.

Fourth Stage - *Peer Review*, an independent scientific review by qualified experts.

Fifth Stage - *Regulatory Acceptance*, adoption for regulatory use by an agency.

The EPA has developed extensive guidance on the conduct of peer reviews because the Agency believes that peer review is an important step in ensuring the quality of science that underlies its regulatory decisions (USEPA (2007)).

Criteria for the validation of alternative test methods *(in vitro* methods designed to replace animal tests in whole or in part) have generally been agreed upon in the United States by the ICCVAM, in Europe by the European Centre for the Validation of Alternative Methods (ECVAM), and internationally by the Organisation for Economic Co-Operation and Development (OECD). These criteria as stated by ICCVAM (NIEHS (1997)) are as follows:

1. The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.
2. The relationship of the endpoints determined by the test method to the in vivo biologic effect and toxicity of interest must be addressed.
3. A formal detailed protocol must be provided and must be available in the public domain. It should be sufficiently detailed to enable the user to adhere to it and should include data analysis and decision criteria.
4. Within-test, intra-laboratory and inter-laboratory variability and how these parameters vary with time should have been evaluated.
5. The test method's performance must have been demonstrated using a series of reference chemicals preferably coded to exclude bias.
6. Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test to that of the test it is designed to replace.
7. The limitations of the test method must be described (e.g., metabolic capability).
8. The data should be obtained in accordance with Good Laboratory Practices (GLPs).
9. All data supporting the assessment of the validity of the test methods including the full data set collected during the validation studies must be publicly available and, preferably, published in an independent, peer-reviewed publication.

The EPA has adopted these various validation criteria for the EDSP as described (USEPA (2007)). Although attempts have been made to thoroughly comply with all validation criteria, the various *in vitro* and *in vivo* screening assays are not replacement assays (Validation Criterion No. 6). Many of them are novel assays; consequently, large data bases do not exist as a reference to establish their predictive capacity (e.g., determination of false positive and false negative rates). It is expected that the review of results from the testing of the first group of 50 to 100 chemicals that was recommended by the Scientific Advisory Panel (SAP) (USEPA (1999)) will allow a more complete assessment of the performance of the Tier-1 screening battery in time.

For technical guidance in developing and validating the various Tier-1 screens and Tier-2 tests, the EPA chartered two federal advisory committees: the Endocrine Disruptor Methods Validation Subcommittee, or EDMVS (from 2001 to 2003), and the Endocrine Disruptor Methods Validation Advisory Committee, or EDMVAC (from 2004 to 2006). These committees, composed of scientists from government, academia, industry, and various interest groups, were charged to provide expert advice to the EPA on protocol development and validation. The EPA also cooperates with member countries of the OECD to develop and validate assays of mutual interest to screen and test for endocrine effects.

It should be remembered that even though assays are being developed and validated individually and peer reviewed on an individual basis (i.e., their strengths and limitations are being evaluated as stand-alone assays), the Tier-1 assays will be used in a battery of complementary screens. An individual assay may serve to strengthen the weight of evidence in a determination (e.g., positive results in an ER binding assay in conjunction with positive results in the uterotropic and pubertal female assays would provide a consistent signal for estrogenicity) or to provide coverage of MOAs not addressed by other assays in the battery. Information supporting the validation of an individual assay may be used at a later date by the Federal Insecticide, Fungicide, Rodenticide Act (FIFRA) SAP for peer review of the EPA's recommendations for a Tier-1 battery. The Tier-1 battery peer review will focus, in part, on the extent of coverage

and overlap the suite of assays will have with one another in detecting endocrine-related effects associated with the EAT hormonal systems.

II. Purpose of this report

The purpose of this Integrated Summary Report is to provide a historical summary of the development and validation of a standardized protocol for the female pubertal rat assay proposed as an *in vivo* assay for the Tier-1 screening battery. The reasoning and judgments leading to the various studies, and conclusions concerning the strengths and weaknesses of the assay in its current form, are presented.

III. Purpose of the assay

The purpose of the female 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 chemicals with antithyroid or estrogenic/anti-estrogenic activity, or agents which alter pubertal development via changes in steroidogenesis, gonadotropin secretion, prolactin, or hypothalamic function.

Weanling rats, standardized to 8 - 10 per litter at post-natal day (PND) 3-5, are housed 2 to 3 per cage. The test chemical is administered in corn oil by oral gavage (2.5 to 5.0 ml/kg) between 0700 and 0900 (lights 14:10, on 0500h) from PND 22 - 42 (21 days) to 15 females per dose level. The endpoints of the assay are shown in Table 3.

Table 3. Endpoints for the female pubertal protocol

	Growth Age at vaginal opening Weight at vaginal opening
Estrous cyclicity	Age at first estrus after vaginal opening Length of estrous cycle Percent of animals cycling Percent of animals cycling regularly
Organ weights	Ovaries (paired) Uterus (blotted) Thyroid Pituitary Adrenals (paired) Liver Kidneys (paired)
Hormones	Serum thyroxine (total) Serum Thyroid Stimulating Hormone
Histopathology	Uterus Thyroid Ovary Kidney and other tissues if indicated

The protocol is included as Appendix 1.

IV. Relevance of the assay

As noted above in Section I.D, relevance describes whether a test is meaningful and useful for a particular purpose. For Tier 1 of the EDSP, the purpose is to identify chemicals with the potential to interact with the endocrine system.

An extensive review of the basis for selecting the endpoints and the conditions of the assay was published (Goldman *et al.* (2000), attached to this Integrated Summary Report as Appendix 2) prior to the validation effort described here, and should be regarded as the primary discussion of the relevance of the endpoints and thus the assay. Because that document, which serves as the Detailed Review Paper (DRP) for the female pubertal protocol, describes in detail the biological relevance of the endpoints, the information presented here will be brief.

As discussed in the female DRP (Goldman *et al.* (2000)), the prepubertal period is a very sensitive age for exposure to agents which alter the endocrine system. The

postweaning period is a time during which a variety of interrelated neuroendocrine processes integrate, and which culminates in the maturation of the reproductive system. The onset of puberty in the female rat is a transitional period which encompasses the period of vaginal opening and first ovulation. Eight days prior to the first ovulation, the first ovulatory cohort of follicles is undergoing growth and they begin to secrete gonadal steroids. Vaginal opening is an estrogen-dependent event that occurs following this growth. For these reasons, compounds which alter neuroendocrine control, steroid production, or bind to the estrogen receptors are able to perturb the onset of puberty.

V. Overview of studies relevant to validation of the assay

The validation process involved 11 positive test chemicals with various modes of action and strengths, and a test chemical which had not previously been tested for endocrine activity but which had been shown to be negative for reproductive and developmental toxicity (Table 4). These studies involved five different contract research laboratories working under Good Laboratory Practices (GLP). The contract studies on the known-positive compounds tested the transferability of the protocol from the developers of the assay (EPA) to contract laboratories.

Following a description of the preliminary contract studies which tested transferability of the assay and examined the applicability across various modes of endocrine action, the results of an interlaboratory comparison study are summarized. This study provided information on the reliability and reproducibility of the assay when conducted in different laboratories.

Results of several in-house studies run by EPA's Office of Research and Development (ORD) laboratories to address specific questions will also be presented in this report. Four chemicals were tested by an ORD laboratory to further assess compounds which alter the hypothalamic-pituitary-gonadal axis and thyroid hormone homeostasis. ORD also conducted a feed restriction study to examine to what extent changes in body weight alone might interfere with interpretation of the endpoints of the assay.

Table 4. Chemicals and doses employed in the validation of the female pubertal protocol

(Dose levels in mg/kg/day. Terminal bodyweight as % of controls shown in parentheses.)

Name of study	TherImmune 1[a]	TherImmune 2	RTI	Interlab Comparison[b]
Attribute of protocol tested	Transferability	Sensitivity	Sensitivity	Reproducibility
Ethynyl estradiol	0.005 *(94,94) (87,89)*	0.0025 *(99)* 0.005 *(100)*	-	-
Propylthiouracil	240 *(68,66) (55,51)*	-	2 *(97)* 25 *(72)*	-
Ketoconazole	100 *(98,95) (91,87)*	-	50 *(99)* 100 *(92)*	-
Pimozide	30 *(88,86) (75,69)*	-	-	-
Methoxychlor	100 *(94,95) (85,89)*	12.5 *(98)* 25 *(100)* 50 *(97)*	25 *(95)* 50 *(94)*	12.5 *(96,100,96)* 50 *(95,94,94)*
Tamoxifen	10 *(82,79) (79,79)*	-	-	-
Phenobarbital	-	25 *(100)* 50 *(100)* 100 *(100)*	-	-
Atrazine	-	-	75 *(92)* 150 *(86)*	-
Fenarimol	-	-	50 *(100)* 250 *(90)*	-
Bisphenol A	-	-	400 *(89)* 600 *(86)*	-
DE-71	-	-	-	30 *(98,100,100)* 60 *(100,100,100)*
2-Chloronitrobenzene	-	-	-	25 *(100,100,100)* 100 *(100,100,97)*

[a] The TherImmune 1 study was conducted in two blocks, in two rat strains each. Terminal bodyweights as percent of control are shown for each group as groups (1,3) (2,4). Groups 1 and 3 were Sprague-Dawley rats, and Groups 2 and 4 were Long Evans rats.

[b] The Interlab Comparison Study was performed in three laboratories. Terminal bodyweights as percent of control are shown for laboratories Argus, WIL, and Huntingdon, respectively.

Transferability of the protocol (TherImmune 1)

A. Purpose

The purpose of this study was to determine if the protocol is transferable. Transferability is the ability of the protocol to be accurately conducted in another laboratory by following the guidance of the assay protocol. The laboratory is assumed to have a reasonable amount of familiarity with reproductive and developmental toxicological techniques, but any specialized techniques necessary to the conduct of the assay must be described sufficiently in the guidance that no further instruction is needed.

The initial study which examined transferability of the assay from the developers to an outside laboratory was the TherImmune single-dose-level study (also referred to as "TherImmune 1" since there was a separate pubertal study, the TherImmune multi-dose-level study, discussed below, which is referred to as "TherImmune 2"). Specific goals of the study were: (1) to assess the transferability of the female protocol (as it existed in 1999), (2) to assess the intra-laboratory variability in endpoint values that might be encountered in a laboratory new to the protocol, and (3) to examine the influence of two different rat strains on the sensitivity of the assay (Long-Evans Hooded and Sprague Dawley). The study was conducted under GLP by an independent, commercial laboratory (TherImmune).

Test chemicals and the dose level of each were selected by the U.S. EPA staff based upon published data demonstrating their ability to alter endocrine function (receptor agonist/antagonist, alter HPG and thyroid homeostasis). Ethynyl estradiol, tamoxifen, propylthiouracil (PTU), ketoconazole, pimozide, and methoxychlor were tested at a single high dose level based on well-known effects in historical studies. The dose levels were selected to maximize the likelihood of demonstrating transferability: if transferability could not be demonstrated with the dose levels chosen, it was thought unlikely that the protocol could be considered transferable for any compound. (See Section X.D for further discussion on dose selection issues.)

The TherImmune 1 study was conducted in two blocks using both Sprague-Dawley and Long Evans rats with six animals/treatment group/block. Separate vehicle

controls (corn oil) were included for each block and strain. Note: this study did not measure thyroid weight, as it was an optional endpoint at that time.

B. Results

The most relevant results of the TherImmune 1 (single dose-level) study are provided in Table 5; the complete results are provided in Appendix 3 (the full report) and Appendix 4 (a summary of all the endpoints)[1].

In general, the data obtained using the protocol successfully demonstrated transferability of the protocol: they identified the expected endocrine-mediated effects on female pubertal development following exposure to chemicals with estrogenic or anti-estrogenic activity, inhibitors of steroid and thyroid hormone synthesis, and a dopamine antagonist. Ethynyl estradiol, tamoxifen (estrogen receptor antagonist and partial estrogen agonist), and methoxychlor (estrogen receptor agonist) advanced the onset of vaginal opening, as expected. On the other hand, ketoconazole (an inhibitor of steroid synthesis) and pimozide (a dopamine antagonist that disrupts the hypothalamic-pituitary control of gonadotropin secretion) delayed the age of vaginal opening, also as expected. In addition to altering the age at vaginal opening, these compounds also influenced the age at first estrus and a number of reproductive endpoints evaluated at necropsy or by histopathology (Table 5). Finally, as predicted, propylthiouracil (an inhibitor of thyroid hormone synthesis) reduced serum T_4 and elevated thyroid stimulating hormone (TSH). It should be noted that changes in serum thyroid hormones were observed after dosing with three other compounds. These changes included an increase in TSH and T_4 with tamoxifen and a decrease with methoxychlor and pimozide. These findings are consistent with other reports in the literature demonstrating both the increase in TSH and T_4 in response to tamoxifen (de Araujo *et al.* (2006)) and decreases in TSH following exposure to methoxychlor (Gray, Jr. *et al.* (1989)). Propylthiouracil at 240 mg/kg/day also delayed vaginal opening in Sprague-Dawley but not Long Evans rats, and this is consistent with what was reported by Marty *et al.* (1999) at the same dose level in the same strain.

[1] Throughout this document, an attempt was made to reduce visual clutter by presenting only the most relevant information. The complete tables with all of the endpoints are included in the Appendices.

Table 5. Summary of significant effects on major endpoints in the transferability study

Treatment	Mode of action	Age at vaginal opening	Age at first estrus	Histopathology[a]	TSH	T_4
Ethynyl estradiol (0.005 mg/kg/d)	ER agonist	↓	↓	Y	-	-
Tamoxifen (10 mg/kg/d)	ER antagonist (partial agonist)	↓	↑	Y	↑	↑
Propylthiouracil 240 mg/kg/day	Inhibits T_3 synthesis	↑SD	↑	Y	↑	↓
Ketoconazole 100 mg/kg/d)	Inhibits steroidogenesis	↑SD	↑	Y	-	↓
Pimozide (30 mg/kg/day)	Dopamine receptor antagonist	↑	-	Y	↓LE	↓
Methoxychlor 100 mg/kg/day)	ER agonist	↓	↓	Y	↓LE	-

Key: ↑ = Significantly increased compared to control LE = Long Evans rats only
↓ = Significantly decreased compared to control SD = Sprague-Dawley rats only
Y = Relevant tissue affected [a]Results are consistent with mode of action

C. Discussion

This study demonstrated that the female pubertal assay is transferable to laboratories which have not had previous experience performing this assay. Results were generally as expected for all the test chemicals. Intra-laboratory variability of organ weights was higher than expected but did not preclude identification of these relatively strong endocrine-active agents as positives. There was generally no outstanding difference between the two rat strains although differences were noted. Finally, the study demonstrated the relevance of the assay to the detection of interaction with the endocrine system through several different mechanisms of action.

As this study was conducted in two complete blocks with a sample size of six animals/treatment/block, there was an opportunity to examine whether a reduced number of animals was sufficient to observe effects reliably. It was found that the small sample size may have limited the detection of significant treatment effects in each block but when the data from the blocks were pooled, the appropriate significant treatment effects were observed.

In addition to demonstrating that the female pubertal protocol was transferable and capable of detecting compounds that interfere with various endocrine modes of action, the results also raised several issues that needed further consideration. One technical issue of note concerned the variability in some tissue weights. This was especially true for weights of tissues with fluid-filled lumina (e.g., uterus) and the smaller tissues such as the adrenal and pituitary glands. For these tissues, the variability appeared to be quite high. To examine this further, the variability for the different endpoints was compared to historical control data on these same endpoints that was available to EPA from other submitted studies. The results of this comparison are shown in Table 6 and Table 7, which compare the coefficients of variation (CVs) for the control data from the "transferability study" with historical control data produced by eight government and commercial laboratories. The CVs for the adrenal and pituitary weights reported by the contractor were 1.5 to 3 times higher than those for historical control data from these sources. Through discussions with the contractor, it was learned that because of the large number of animals killed at each necropsy, there were delays in weighing tissues. Thus, some of the smaller tissues and those containing fluid may have partially dried prior to weighing. This led to improvement of the protocol for subsequent studies by including specific instructions to avoid dessication of small organs between necropsy and weighing.

It is also important to note that the variation in the uterine weights was expected, since uterine weights fluctuate during the estrous cycle and these females were killed on various days of their cycles.

Table 6. Comparison of the transferability study with historical data for age at vaginal opening in control Sprague-Dawley, Wistar and Long Evans rats

	Historical data							Transfer-ability study
	Contract Lab 1	Contract Lab 2	Contract Lab 3	Contract Lab 4	Contract Lab 5	Contract Lab 6	EPA Lab 1	
Sprague-Dawley	30.9 ± 1.5	31.4 ± 1.1	31.8 ± 0.69	32.0 ± 1.5	33.6 ± 2.3	33.1 ± 2.6	32.7 ± 1.1	34.9 ± 1.3
	CV= 4.85%	3.50%	2.17%	4.69%	6.85%	7.85%	3.36%	3.75%
Wistar	34.3 ± 1.3	34.1 ± 1.3	33.7 ± 0.97	35.4 ± 1.8	32.8 ± 2.0	32.4 ± 2.1	33.1 ± 1.2	
	CV= 3.79%	3.82%	2.88%	5.08%	6.14%	6.38%	3.74%	
	EPA Lab 2	EPA Lab 2	EPA Lab 2	EPA Lab 2				Transfer-ability study
Long Evans	33.4 ± 1.8	32.25 ± 2.0	33.7 ± 1.5	30.6 ± 1.2				36.3 ± 2.1
	CV= 5.39%	6.36%	4.39%	3.89%				5.67%

Data are reported as mean ± standard deviation; CV= coefficient of variation

Table 7. Comparison of TherImmune data with historical data for necropsy body weight and tissue weights in control Sprague-Dawley, Wistar and Long Evans female rats

	Historical data					Transferability study	
	Contract Lab 5	Contract Lab 5	Contract Lab 2	EPA Lab 2	EPA Lab 2	TherImmune	TherImmune
	Sprague-Dawley	Sprague-Dawley	Wistar	Wistar	Wistar	Sprague-Dawley	Long Evans
Body weight (g)		387 ± 30	150 ± 8.7	140 ± 7.7	303 ± 20	146 ± 8.0	161 ± 8.8
		CV= 7.75%	CV= 5.79%	CV= 5.49%	CV= 6.60%	CV= 5.45%	CV= 5.49%
Liver (g)	12.2 ± 1.54	17.6 ± 2.3	6.73 ± 0.80	6.09 ± 0.53		6.06 ± 0.45	6.96 ± 0.72
	12.6%	13.1%	11.8%	8.74%		7.43%	10.3%
Kidney (g)	2.36 ± 0.29	2.73 ± 0.27	1.42 ± 0.12	1.33 ± 0.09		1.23 ± 0.09	1.45 ± 0.13
	12.2%	9.89%	8.58%	7.32%		7.32%	8.97%
Uterus + fluid (g)		0.53 ± 0.09*	0.354± 0.205	0.346± 0.231	0.525 ± 0.09*	0.29 ± 0.15	0.27 ± 0.11
		16.9%	57.9%	66.8%	17.7%	51.7%	40.7%
Ovary (g)	0.141 ± 0.019	0.109 ± 0.017	0.072 ± 0.019	0.072 ± 0.019	0.061 ± 0.011	0.065 ± 0.013	0.082± 0.017
	13.5%	15.6%	26.4%	26.4%	18.0%	20.0%	20.7%
Adrenals (g)		0.066 ± 0.008	0.045 ± 0.005	0.038 ± 0.005		0.043 ± 0.008	0.037 ± 0.007
		12.1%	11.1%	13.2%		18.6%	18.9%
Pituitary (g)	0.016 ± 0.003	0.013 ± 0.020	0.0079±0.0008	0.0076±0.0007		0.007 ± 0.002	0.007 ± 0.003
	18.7%	15.4%	10.1%	9.2%		28.6%	42.9%

Note: These data are presented for comparison of the coefficients of variation, which is still possible even though the ages of females at necropsy varied between studies and thus the absolute values vary.
Data are reported as mean ± standard deviation; CV= coefficient of variation; * All females in these groups were killed during diestrus.

By evaluating two strains of rats in these studies, it was possible to determine if the expected endocrine-mediated changes in pubertal development could be detected in Sprague-Dawley and Long Evans rats. Again, the results depicted in Table 5 demonstrate that although there are some discrepancies, the two strains responded in a very similar pattern, arguing that, in this case, the two strains were not very discrepant. Importantly, subsequent studies using still other rat strains (e.g., Wistar (Laws *et al.* (2000a)) and Alpk:ApfSD (Ashby *et al.* (2002))) have demonstrated that the female pubertal protocol can be employed successfully in a variety of strains. Strain differences, or lack thereof, are discussed further in Section X.B.

Feedback from the contractor indicated other areas in the protocol where clarity and more-detailed technical direction would have been helpful. As a result, the protocols were edited to insure that each key step is clearly described. For example, the statistical analysis section required editing to better describe the options for data analysis.

This transferability study was conducted using high doses of potent chemicals. Following this study, it was recommended that the sensitivity of the pubertal protocols for weaker endocrine-active chemicals and lower doses be explored, to define how robust the protocols will be as a screen for identifying endocrine mediated effects.

In addition, because of the potential impact of reduced body weight on the endpoints measured in the female pubertal protocol and the practice of using a 10% decrease in body weight when setting the maximum tolerated dose (MTD), a study to characterize the effects of reduced food intake and body weight on the endpoints was recommended.

Both suggestions were acted on by the Agency and the results of these studies are discussed below in Section VI.

VI. Sensitivity of the protocol

Two studies were undertaken to demonstrate the sensitivity of the female pubertal protocol. One study, referred to as the "multi-chemical study", focused on further examining the ability of the assay to respond to chemicals with different modes

of interaction with the endocrine system, using weaker chemicals or lower dose levels than were used in the initial transferability study. The other study, referred to as the "multi-dose study", examined the response of the assay to dose levels near the lowest observed adverse effect level for a weak estrogen and a weak thyroid-active agent. These studies overlapped on one chemical and included several chemicals that were tested in the transferability study (Table 4), to provide an initial evaluation of the reproducibility of results across laboratories.

A. Multi-chemical study (RTI)

1. Purpose

The purpose of this study, conducted by RTI International, was to examine the response of the pubertal assay to the effects of a wide variety of chemicals that are known to affect the endocrine system through different pathways and/or mechanisms of action, and to obtain a sense of the ability of the assay to detect weaker chemicals. A detailed report of this exercise is presented in Appendix 5[2], and a summary of the results for all of the endpoints is presented in Appendix 6.

Six chemicals representing different modes of action were selected by EPA and the experiments were performed in a commercial laboratory. The study was conducted in two components. Component 1 consisted of animals dosed with atrazine (75 or 150 mg/kg/day), fenarimol (50 or 250 mg/kg/day), or methoxychlor (25 or 50 mg/kg/day), and a concurrent vehicle control (corn oil) group. Component 2 consisted of animals receiving bisphenol A (400 or 600 mg/kg/day), ketoconazole (50 or 100 mg/kg/day), or propylthiouracil (2 or 25 mg/kg/day), and a concurrent vehicle control (corn oil) group. The six chemicals and their modes of action are shown in Table 8.

[2] The final report was written by the contractor based partly on analyses of covariance (ANCOVA) with *terminal* body weight, as requested by EPA. The analyses presented in this Integrated Summary Report, however, are based on analyses of convariance with *weaning* body weight. There may be therefore be differences between the conclusions in the contractor's final report and the conclusions presented here. For a discussion of why ANCOVA with body weight at weaning is preferred, see Section X.E. The ANCOVA analyses using body weight at weaning as covariate are attached as Appendix 7.

Table 8. Compounds and doses selected for the multi-chemical (RTI) study

Compound	Dose (mg/kg/day)	Mode of action
Atrazine	75 150	CNS/pituitary toxicant
Fenarimol	50 250	Weak aromatase inhibition
Methoxychlor	25 50	Estrogen agonist
Bisphenol A	400 600	Estrogen agonist
Ketoconazole	50 100	Inhibits steroidogenesis
Propylthiouracil	2 25	Inhibits T_3, T_4 synthesis

The doses were selected in an attempt to approximate the Maximum Tolerated Dose (MTD, defined as the dose causing approximately a 10% decrease in body weight compared to controls by the end of the exposure period), and ½ MTD. In some cases, however, such as propylthiouracil (PTU) and bisphenol A (BPA), MTD for juvenile/pubertal animals was not known and dose level selection was based on the best available information. For example, the dose level for PTU in the transferability study (240 mg/kg/day) led to terminal body weights that were only 40% of controls so the dose levels in this study were reduced drastically. There was also uncertainty about appropriate dose levels for bisphenol A. Laws *et al.* (2000a) reported no decrease in age at vaginal opening from BPA at 400 mg/kg/day in Long Evans rats, while Tyl *et al.* (2002) reported that a concentration in feed of 7500 ppm (corresponding to an estimated 500 mg/kg/day) exceeded the MTD in adults in a multi-generation reproductive toxicity study and delayed vaginal opening (rather than accelerating it). It was not clear whether the high dose level chosen for this study, 600 mg/kg/day, would be too high, too low, or close to the MTD, as targeted.

2. Results

a. Atrazine

Atrazine is a chlorinated triazine herbicide used on grasses and weeds. Although the primary cellular mechanism for this compound's effects on endocrine function have not been fully characterized, it is well established that there is a disruption of the hypothalamic (central nervous system, CNS) control of pituitary function. In particular, the ovulatory surge of luteinizing hormone is decreased in a dose-dependent manner, resulting in a similar dose-dependent disruption of the female estrous cycle (Cooper *et al.* (1996);Cooper *et al.* (2000);McMullin *et al.* (2004)). This mode of action was also implicated in a delay in puberty following exposure to atrazine and its primary metabolites using both male and female Wistar rats (Stoker *et al.* (2000);Stoker *et al.* (2002);Laws *et al.* (2000b);Laws *et al.* (2003);Ashby *et al.* (2002)). Thus, atrazine was used in the present study employing Sprague-Dawley rats to confirm that this compound, which alters CNS control of pituitary-gonadal function, would alter pubertal development in this strain when tested in a contract laboratory.

Exposure to atrazine at 75 or 150 mg/kg/day delayed puberty, as evidenced by delayed vaginal opening (Table 9). More specifically, the postnatal day of vaginal opening exhibited a significant delay at the high dose (36.4 days vs. 32.9 days for the control group) with a significant trend for a delay when evaluating both doses. Body weight and body weight gain were statistically decreased compared to controls at both dose levels at the terminal kill (93.3 and 87.8% for the 75 and 150 mg/kg dose levels respectively), although not at acquisition of vaginal opening. Ovarian weight was significantly decreased in the highest dose. There was no effect on uterine weight. Pituitary weight was decreased at the high dose. Circulating T_4 or TSH levels, or histopathological changes in the thyroid, ovaries, or uterus were not affected by treatment, and the day of first estrus was also significantly delayed. The delay in vaginal opening (VO, in the absence of any significant change in body weight on the day of VO), the delay in first estrus (described in Appendix 5), and the impaired vaginal cycling following the first estrus (also described in Appendix 5) are in agreement with the previous atrazine studies by Laws *et al.* (2000b) and Laws *et al.* (2003) using Wistar

rats, and with Ashby *et al.* (2002) using both Sprague-Dawley and Alpk:ApfSD (Wistar-derived) rats. The NOAEL in this study was 75 mg/kg. Laws *et al.* (2000b) and Ashby *et al.* (2002) reported a delay in VO at 50 and 30 mg/kg/day (Sprague-Dawley rats) respectively.

The conclusion from this part of the study was that the female pubertal assay detected interaction with the endocrine system from a compound which affects the HPG axis, when tested in a contract laboratory using the Sprague-Dawley strain of rats.

Table 9. Effect of atrazine on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	75	150
Age at VO (PND) [b]	32.9 ± 0.4	34.2 ± 0.4	36.4 ± 0.6 *†
BW at VO (g)	127.89 ± 3.20	128.1 ± 3.03	129.81 ± 2.91
Initial BW (g)	58.42 ± 0.97	58.91 ± 1.16	58.55 ± 1.22
Final BW (g) [b]	183.39 ± 3.61	171.02 ± 2.73 *†	160.98 ± 3.06 *†
Final BW as % of control		93.3	87.8
BW gain [b]	124.97 ± 3.00	112.12 ± 2.45 *	102.43 ± 2.29 *
Pituitary wt (mg)	10.5 ± 0.4	9.9 ± 0.7	8.3 ± 0.3 *†
Thyroid wt (mg)	19.3 ± 2.1	16.2 ± 1.0	16.1 ± 1.2
Liver wt (g)	9.71 ± 0.37	9.0 ± 0.21	8.7 ± 0.36 †
Adrenal wt, paired (mg)	43.2 ± 2.7	47.8 ± 2.2	41.4 ± 2.1
Kidney wt, paired (g)	1.86 ± 0.049	1.81 ± 3.55	1.74 ± 0.05
Ovary wt, paired (mg)	101.5 ± 4.6	96.2 ± 4.0	82.3 ± 4.4 *†
Uterus wt w/fluid (mg)	369.4 ± 28.5	356.8 ± 34.5	321.5 ± 25.1
Uterus wt wo/fluid (mg)	343.3 ± 18.6	329.9 ± 22.3	296.5 ± 22.1
Thyroxine (ug/dl)	4.44 ± 0.24	4.55 ± 0.21	4.67 ± 0.24
TSH (ng/ml)	9.62 ± 0.53	9.9 ± 0.79	8.52 ± 0.63
Diestrus	2	2	3
Proestrus/Estrus	6	7	6
Not cycling	2	1	1

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
Values for estrous cycle stage represent the number of animals in that stage at kill.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

b. Fenarimol

Fenarimol was selected as a weak aromatase inhibitor based on reports in the literature examining this fungicide *in vivo (*Hirsch *et al.* (1987a);Andersen *et al.* (2002);Hirsch *et al.* (1987b);Hirsch *et al.* (1986)).

In this study, fenarimol at 50 and 250 mg/kg/day did not alter the day of acquisition of vaginal opening compared to the control group (Table 10). However, dry uterine weight and pituitary weight were significantly lower than controls at the high dose. The ovarian weight was reduced, although this effect was present at $p = 0.06$, marginally beyond the $p < 0.05$ cut-off for statistical difference. The changes noted in the reproductive tissue weights would be consistent with a fenarimol-induced decreased in estrogen biosynthesis. The lack of effect on VO is not what would be expected if this compound decreases serum estrogens.

It should be noted that Marty *et al.* (1999) reported that the more potent and specific aromatase inhibitor fadrozole affected a number of endpoints in the female pubertal protocol. Specifically, they reported delayed age at vaginal opening and increased body weight at vaginal opening, as well as decreased uterine weight, at all three dose levels tested (0.6, 1.2, and 6.0 mg/kg/day). Thus, there is reason to believe that the female pubertal assay is sensitive to aromatase inhibitors, even though there was little response to fenarimol in this multi-chemical study. Fadrozole was the first choice for studying an aromatase inhibitor in this multi-chemical study, but it could not be obtained.

Recent studies have shown that fenarimol possesses the ability to affect estrogen and androgen receptor function in addition to inhibiting aromatase. For example, significant estrogenic activity (i.e., increased uterine weight at 200 mg/kg for 4 days in the uterotrophic assay, Andersen *et al.* (2006)) and anti-androgenic activity (i.e., decreased prostate weight in the Hershberger assay at 200 mg/kg, Vinggaard *et al.* (2005)) have been reported. Thus, in retrospect, fenarimol has mixed modes of action at the dose levels used in this study that may explain the complex response observed. Nevertheless, the overall action of this compound would have raised concern for potential interaction with the endocrine system.

There was a statistically significant decrease in thyroid weight at both dose levels when adjusted for covariance with body weight at weaning. In agreement with other reports (Andersen *et al.* (2006)), there were changes in thyroid hormone levels. An increase in circulating TSH levels was observed at 50 and 250 mg/kg/day. A decrease in the circulating T_4 level was observed at the high dose. These effects on thyroid hormone levels were consistent with the observation of follicular hypertrophy in 33% of the high-dose thyroids.

This part of the study showed that the female pubertal assay does detect changes in endocrine function induced by fenarimol when tested in a contract laboratory using Sprague-Dawley rats. However, in retrospect, the choice of fenarimol to demonstrate sensitivity of the assay to aromatase inhibitors was unfortunate since studies published subsequent to this one showed that fenarimol also possesses estrogenic activity. It is therefore not surprising that the effects of fenarimol on the female pubertal assay's endpoints were difficult to interpret, even though there were indications of impaired estrogen activity in uterine weight, ovarian weight, and body weight at vaginal opening. Fenarimol would, however, have been identified as altering thyroid function, and this highlights the ability of the female pubertal protocol to detect chemicals that act on different parts of the endocrine system.

Table 10. Effect of fenarimol on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	50	250
Age at VO (PND)	32.9 ± 0.4	33.7 ± 0.4	34.2 ± 0.7
BW at VO (g) [b]	127.89 ± 3.20	132.62 ± 2.45	114.53 ± 4.35 *†
Initial BW (g)	58.42 ± 0.97	59.44 ± 1.06	58.56 ± 1.43
Final BW (g) [b]	183.39 ± 3.61	186.32 ± 2.93	166.48 ± 3.51 *†
Final BW as % of control		101.6	90.8
BW gain	124.97 ± 3.00	126.88 ± 2.77	107.92 ± 2.81 *
Pituitary wt (mg)	10.5 ± 0.4	10.5 ± 0.5	8.7 ± 0.2 *†
Thyroid wt (mg)	19.3 ± 2.1	15.4 ± 1.1 †	15.1 ± 0.9 †
Liver wt (g)	9.71 ± 0.37	11.48 ± 0.38 †	14.19 ± 0.90 *†
Adrenal wt, paired (mg)	43.2 ± 2.7	52.4 ± 3.0 †	44.1 ± 2.8
Kidney wt, paired (g)	1.86 ± 0.049	1.97 ± 0.053	1.76 ± 0.044
Ovary wt, paired (mg)	101.5 ± 4.6	110.6± 5.1	89.5 ± 4.4
Uterus wt w/fluid (mg)	369.4 ± 28.5	372.9 ± 25.5	311.8 ± 31.7
Uterus wt wo/fluid (mg)	343.3 ± 18.6	352.6 ± 20.1	278.2 ± 18.7 *†
Thyroxine (ug/dl)	4.44 ± 0.24	4.72 ± 0.24	3.73 ± 0.13 *
TSH (ng/ml)	9.62 ± 0.53	12.53 ± 0.81 *	19.65 ± 2.29 *
Diestrus	2	3	4
Proestrus/Estrus	6	7	5
Not cycling	2	2	3

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

c. Methoxychlor

Methoxychlor is an organochlorine insecticide effective against a wide range of pests encountered in agriculture, households, and ornamental plantings. It was registered for use on fruits, vegetables, forage crops, and in forestry. Methoxychlor is also registered for veterinary use to kill parasites on dairy and beef cattle. Available evidence suggests that high doses of technical methoxychlor (88 to 90% pure) or its metabolites may have estrogenic or reproductive effects (Smith (1991)). Laws *et al.* (2000a) showed that exposure to 50 mg/kg of methoxychlor from PND 21 to 35 advanced VO in rat.

The results of the current study on methoxychlor is shown in Table 11.

Table 11. Effect of methoxychlor on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	25	50
Age at VO (PND) [b]	32.9 ± 0.4	29.8 ± 0.6 *†	27.4 ± 0.3 *†
BW at VO (g) [b]	127.89 ± 3.20	104.85 ± 4.13 *†	89.54 ± 2.40 *†
Initial BW (g)	58.42 ± 0.97	58.73 ± 1.08	58.44 ± 1.21
Final BW (g) [b]	183.39 ± 3.61	174.38 ± 2.87†	171.62 ± 2.59 *†
Final BW as % of control		95.1	93.6
BW gain [b]	124.97 ± 3.00	115.57 ± 2.17 *	113.18 ± 2.52 *
Pituitary wt (mg)	10.5 ± 0.4	9.7 ± 0.4	9.4 ± 0.51
Thyroid wt (mg)	19.3 ± 2.1	14.9± 0.7 †	16.8 ± 0.8
Liver wt (g)	9.71 ± 0.37	8.86 ± 0.28	8.85 ± 0.25
Adrenal wt, paired (mg)	43.2 ± 2.7	43.9 ± 3.0	46.3 ± 3.0
Kidney wt, paired (g)	1.86 ± 0.049	1.82 ± 0.051	1.75 ± 0.051
Ovary wt, paired (mg)	101.5 ± 4.6	91.6 ± 4.1	86.7 ± 4.9 †
Uterus wt w/fluid (mg)	369.4 ± 28.5	409.2 ± 46.7	464.5 ± 61.5
Uterus wt wo/fluid (mg)	343.3 ± 18.6	350.4 ± 24.3	365.2 ± 26.7
Thyroxine (ug/dl)	4.44 ± 0.24	4.57 ± 0.15	4.69 ± 0.17
TSH (ng/ml)	9.62 ± 0.53	8.34 ± 0.47	9.83 ± 0.98
Diestrus	2	2	3
Proestrus/Estrus	6	8.	7
Not cycling	2	0	3

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

In this study, female rats were exposed to 25 and 50 mg/kg methoxychlor. Both dose levels decreased body weight, though neither dose reached the MTD, when weights were adjusted for covariance with body weight at weaning (Table 11). Vaginal opening was accelerated at both 25 and 50 mg/kg/day (PND 29.6 and 27.4 respectively, compared to 32.9 for controls), consistent with Laws *et al.* (2000a); and body weight at acquisition was decreased compared to the control. Decreased adjusted ovary weight was observed at the high dose, as expected for an estrogenic chemical (Yamasaki *et al.* (2002a);Yamasaki *et al.* (2002b)). There was a nonsignificant increase in uterine wet weight ($p = 0.08$ adjusted for weaning weight), and a non-significant decrease in pituitary

weight (p=0.10). This result in pituitary weight is consistent with Ostby *et al.* (1999), who showed that while estradiol increased pituitary weight, methoxychlor at 200 mg/kg/day led to a non-significant decrease. There was no treatment effect on other adjusted organ weights or histopathology.

This part of the study showed that methoxychlor was clearly identified as interacting with the endocrine system using the female pubertal assay when run at dose levels lower than that used in the transferability study. This confirms transferability of the assay and provides evidence that the assay is sensitive: vaginal opening was the most sensitive indicator, occurring at the low dose in the absence of other effects. Effects at the high dose were consistent with the acceleration of VO at the low dose, and occurred in the absence of indications of other toxicity.

d. Bisphenol A

Bisphenol A (BPA) is a key building block of polycarbonate plastic. It was first reported to be estrogenic in 1936, in female rats (Dodds *et al.* (1936)). More recently, it was found to be estrogenic in the MCF-7 human breast cancer cell culture assay (Krishnan *et al.* (1993)). The hormonal effects could be measured at concentrations as low as 2-5 ppb (2-5 µg/l), suggesting that BPA is fairly potent *in vitro*. Tyl *et al.* (2002), however, examined the effects of bisphenol A in an *in vivo,* three generation study and found limited effects on female reproductive development even at high dose levels, thus suggesting that the dose levels in the pubertal study should be set at a high level.

In the Tyl *et al.* (2002) study, BPA was administered by diet at several doses ranging from 0.015 to 7500 ppm, corresponding to approximately 0.001 to 500 mg/kg/day. The highest dose group consumed approximately 80-100 mg/kg/day during the period of exposure used in the pubertal assay. Tyl *et al.* (2002) found adult systemic toxicity at 750 and 7500 ppm (the two highest doses used) in all generations. These effects included reduced body weight, reduced body weight gain, reduced absolute and increased relative weanling and adult organ weights (liver, kidneys, adrenals, spleen, pituitary and brain) and, in the female, renal and hepatic pathology at 7500 ppm. Importantly, these high dietary doses were without effect on reproductive organ histopathology and function. Ovarian weights, as well as the total number of

pups and live pups/litter on PND 0, were decreased at 7500 ppm, which exceeded the adult maximum tolerated dose (based on body weight, liver and kidney toxicity). At 7500 ppm, vaginal opening was delayed in the offspring and the body weight of these females was 90% of control animal weights. Furthermore, the delay in vaginal opening would not be expected considering the purported estrogenic activity of BPA. Anogenital distance (AGD) on PND 0 was altered in the F2, but not F3 females. Thus, in the multigenerational feeding study, the effects of bisphenol A on female reproductive development were limited.

Dose selection in this pubertal study was based on observations by Laws *et al.* (2000a) who reported a uterotrophic effect of BPA following oral gavage for three days at 200 mg/kg/day in Long Evans rats, but that this uterine weight increase was markedly less than observed following subcutaneous exposure for three days in prepubertal females. Laws *et al.* (2000a) also reported that dose levels as high as 400 mg/kg/day failed to advance the age at vaginal opening. Age at vaginal opening is the primary endpoint in the female pubertal assay. Thus there was considerable uncertainty over whether the dose levels chosen for this study (400 and 600 mg/kg/day) were appropriate.

In this multi-chemical study, age at vaginal opening was not advanced, and body weight at vaginal opening was increased at the high dose (Table 12). This is not what would be expected of an estrogen.

The low and high doses resulted in an 11.3 and 14.2% reduction in body weight, respectively, at the time of kill. This decrease in body weight is at the margin of acceptability since body weight decreases of this magnitude may interfere with other endocrine endpoints of the study (see discussion of feeding study, Section VI.C.2). However, estrogens themselves lead to reduction in body weight gain so this reduction could instead be interpreted as evidence of estrogenicity.

Bisphenol A decreased ovarian weight at both doses. This is in agreement with Tyl *et al.* (2002) and is typical of an estrogenic chemical (Yamasaki *et al.* (2002a);Yamasaki *et al.* (2002b)). Uterine weight (wet and blotted) was also lower in the high dose of BPA. The decrease in uterine weight is different from the increase

observed in the uterotrophic assay. However, this difference is not surprising given the variability of this endpoint in the intact cycling female compared to that of the ovariectomized or pre-pubertal female used in the uterotrophic assays (Laws *et al.* (2000a);Kim *et al.* (2005)).

The status of the ovarian cycle at the time of kill is also difficult to interpret in this study. A review of the vaginal smear data, on which the cycling status of the female is based, reveals that very few of the animals in either the control or treated groups were cycling normally. In this exercise, there were two control groups as the study was conducted in two blocks. A reinspection of the data revealed that regular cycling could be identified in 11 of 13 rats in the control group run concomitantly with the first three chemicals tested (atrazine, fenarimol and methoxychlor), demonstrating that it is reasonable to assume that regular cycling can be established within the short time span of this assay. However, a regular cycling pattern could be identified in only two of 14 animals in the control group used as the basis of comparison for bisphenol A. Furthermore, an ovarian cycle could be discerned for only 6 of 11 and 5 of 14 rats in the low and high dose of bisphenol A, respectively. Thus, whether or not bisphenol A altered ovarian cycles and whether this alteration may have contributed to the differences in tissue weights observed in this study requires further investigation.

In summary, this study did not detect the expected estrogenicity of bisphenol A. This finding is best explained by the fact that oral administration of bisphenol A requires markedly greater doses than subcutaneous injections. Ashby *et al.* (1998) demonstrated that 4 of 7 females showed advanced VO when dosed with 600 mg/kg subcutaneously compared to 0 of 7 females receiving the same dose orally. This route-specific response is likely due to pharmocodynamic differences of bisphenol A when using different routes of exposure.

The results, however, are consistent with what would be expected from general toxicity: decreased terminal body weight, and decreases in liver, kidney, and ovary weights. The negative results for endocrine interaction in this part of the study suggest that the female pubertal assay does not respond to general toxicity, which is appropriate for an assay that is intended to be specific to endocrine interaction.

Table 12. Effect of bisphenol A on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	400	600
Age at VO (PND)	32.3 ± 0.4	33.0 ± 0.6	32.3 ± 0.7
BW at VO (g)	122.70 ± 3.17	116.36 ± 3.5	108.42 ± 4.0 *†
Initial BW (g)	59.20 ± 0.81	58.48 ± 0.97	59.32 ± 0.82
Final BW (g)	180.68 ± 2.84	160.21 ± 2.38 *†	155.05 ± 2.91 *†
Final BW as % of control		88.7	85.8
BW gain	121.11 ± 2.79	101.82 ± 1.45 *	95.86 ± 2.50 *
Pituitary wt (mg)	8.7 ± 0.9	7.6 ± 0.6	7.5 ± 0.7
Thyroid wt (mg)	14.3 ± 0.8	12.6 ± 0.5	13.4 ± 0.7
Liver wt (g)	9.75 ± 0.31	8.11 ± 0.30 *†	7.85 ± 0.29 *†
Adrenal wt, paired (mg)	47.8 ± 2.9	41.4 ± 1.9	41.2 ± 1.6
Kidney wt, paired (g)	1.81 ± 0.053	1.63 ± 0.043 *†	1.59 ± 0.049 *†
Ovary wt, paired (mg)	100.9 ± 4.8	73.4 ± 5.0 *†	66.3 ± 4.9 *†
Uterus wt w/fluid (mg)	327.6 ± 26.1	269.5 ± 21.2	243.6 ± 29.3 †
Uterus wt wo/fluid (mg)	305.7 ± 25.5	253.2 ± 21.4	208.0 ± 20.3 *†
Thyroxine (ug/dl)	4.81 ± 0.27	4.88 ± 0.24	5.23 ± 0.24
TSH (ng/ml)	7.88 ± 0.41	6.97 ± 0.30	7.51 ± 0.42
Diestrus	6	4	3
Proestrus/Estrus	4	3	4
Not cycling	3	1	3

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

e. Ketoconazole

Ketoconazole is a fungicide with well-documented effects on steroidogenesis and the compound was used in this exercise in order to determine whether the female pubertal protocol's endpoints were sensitive to such disruption.

Ketoconazole decreased final body weight and the rate of body weight gain at the high dose. However, this effect did not surpass 10%, indicating that the Maximum Tolerated Dose was not exceeded. VO in the high dose was delayed slightly, but the difference ($p = 0.07$) was not statistically different from controls (Table 13). Body weight

at VO was not affected but ovarian weight and uterine weight (both with and without fluid) decreased.

In addition, ovarian pathology was noted that increased in incidence and severity with dose. This included the presence of cytoplasmic vacuolization of the corpora lutea observed in 12/15 and 9/15 of the low and high-dose animals, respectively. The five remaining animals in the high-dose group exhibited a complete absence of corpora lutea.

In an attempt to determine whether the observed ovarian histology effects could be attributed to some change in ovarian cycles, the vaginal smear data were re-examined. Although few animals in either group were found to exhibit regular ovarian cycles (2 of 9 in the low dose and 3 of 15 in the high dose), the lack of regular cycles in the concurrent controls precluded any conclusive statements as to whether or not this chemical affected cycling and in a way that could explain the altered histology.

Adrenal weight increased markedly at both the low and high doses, consistent with inhibition of steroidogenesis (Creange *et al.* (1978)).

Thus, the clear alterations in ovarian weight, ovarian histopathology, and uterine weight (both wet and dry weight) identify ketoconazole as having the potential to interact with the endocrine system.

The organ weight findings are in good agreement with those reported by Marty *et al.* (1999) who, using three doses of ketoconazole (24, 50, and 100 mg/kg/day) in Sprague-Dawley (CD) rats (the same strain used in this multi-chemical study), found a dose-dependent decrease in absolute and relative uterine weights at 50 and 100 mg/kg and a 60 and 70% reduction in ovarian weight at these two doses, respectively. (Marty *et al.* (1999) did not adjust organ weights for covariance with body weight at weaning.) Again, although the day of vaginal opening was delayed at the high dose in this multichemical study (PND 32.3 in controls vs. 33.7 at the high dose), the effect was not significant. Marty *et al.* (1999), using the same doses, found a significant delay in VO at the 100 mg/kg/day dose. These authors did not perform any reproductive tissue histology.

This part of the study confirmed that the female pubertal assay is capable of identifying this steroidogenesis inhibitor as interacting with the endocrine system when tested in a contract laboratory. The same dose level (100 mg/kg/day) that was associated with effects in the single-dose transferability study was identified with effects in this study, although the effects were not identical. It was notable that adrenal weights responded with a significant increase at the low dose for this steroidogenesis inhibitor, but there were no other clearly identifiable endocrine responses at the low dose, indicating either that the limit of sensitivity of the assay had been reached or that there are no effects at this dose level.

Table 13. Effect of ketoconazole on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	50	100
Age at VO (PND)	32.3 ± 0.4	33.0 ± 0.4	33.7 ± 0.6
BW at VO (g)	122.70 ± 3.17	126.54 ± 2.54	122.6 ± 3.65
Initial BW (g)	59.20 ± 0.81	59.58 ± 1.21	58.96 ± 0.96
Final BW (g)	180.68 ± 2.84	181.12 ± 2.99	166.39 ± 2.96 *†
Final BW as % of control		100.2	92.1
BW gain	121.11 ± 2.79	121.11± 2.11	107.42 ± 2.54 *
Pituitary wt (mg)	8.7 ± 0.9	8.8 ± 0.5	8.0 ± 0.4
Thyroid wt (mg)	14.3 ± 0.8	14.7± 0.5	14.4 ± 0.7
Liver wt (g)	9.75 ± 0.31	10.77 ± 0.30 †	10.56 ± 0.40 †
Adrenal wt, paired (mg)	47.8 ± 2.9	79.9 ± 3.9 *†	78.5 ± 3.7 *†
Kidney wt, paired (g)	1.81 ± 0.053	1.96 ± 0.052	1.98 ± 0.05 †
Ovary wt, paired (mg)	100.9 ± 4.8	96.8 ± 5.1	84.3 ± 3.7 *†
Uterus wt w/fluid (mg)	327.6 ± 26.1	298.1 ± 23.9	239.0 ± 15.8 *†
Uterus wt wo/fluid (mg)	305.7 ± 25.5	273.0 ± 17.2	221.7 ± 15.2 *†
Thyroxine (ug/dl)	4.81 ± 0.27	4.97 ± 0.16	4.79 ± 0.19
TSH (ng/ml)	7.88 ± 0.41	8.91 ± 0.82	8.24 ± 0.47
Diestrus	6	5	6
Proestrus/Estrus	4	5	2
Not cycling	3	1	0

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

f. Propylthiouracil

Propylthiouracil is a well known thyrotoxicant, targeting the uptake of iodine into the gland and thereby inhibiting the synthesis of T_3 and T_4. As expected, both doses of PTU selected for this study led to a significant decrease in serum T_4 and a consequent increase in blood levels of TSH. This compound also resulted in an increase in thyroid weight (Table 14), and thyroid follicular cell hypertrophy/hyperplasia at both 2 and 25 mg/kg/day. It is important to note that the 2 mg/kg/day dose significantly altered the thyroid hormone measures without affecting body weight. Vaginal opening was unaltered by propylthiouracil treatment and this compound was without effect on any of the other reproductive endpoints.

Results at the high dose, while consistent with results at the low dose and with the single-dose transferability study, must be interpreted with caution due to the high (28%) body weight loss compared to controls. See Section VI.C.3 for further discussion of the effect of body weight on thyroid and other endpoints.

The observed decrease in T_4 and increase in TSH following 2 mg/kg/day of PTU is similar to that reported more recently in a study by Yamasaki *et al.* (2002c), in which a dose of 1 mg/kg/day to the *male* rat from PND23 – PND53 (i.e., longer than the PND 22 - PND 42 exposure to the females in this study) resulted in a reduction in T_4 and an increase in TSH. They identified 0.1 mg/kg/day as the NOEL for this compound.

This part of the multi-chemical study showed that the female pubertal assay is clearly capable of identifying interaction of a chemical with the thyroid system at a dose which does not cause a change in body weight or other signs of general toxicity. In addition, a dose was identified that, although potent with respect to the thyroid, was without effect on the reproductive axis.

Table 14. Effect of propylthiouracil on pubertal female endpoints

	Dose level (mg/kg/day)		
	0	2	25
Age at VO (PND)	32.3 ± 0.4	33.3 ± 0.5	33.1 ± 0.4
BW at VO (g) [b]	122.70 ± 3.17	127.35 ± 2.53	117.71± 1.86
Initial BW (g)	59.20 ± 0.81	58.84 ± 0.94	59.54 ± 0.87
Final BW (g) [b]	180.68 ± 2.84	175.77 ± 2.73	130.75 ± 2.57 *
Final BW as % of control		97.3	72.4
BW gain [b]	121.11 ± 2.79	116.93 ± 2.82	71.21 ± 2.18 *
Pituitary wt (mg)	8.7 ± 0.9	8.5 ± 0.6	7.5 ± 0.5
Thyroid wt (mg)	14.3 ± 0.8	38.5± 1.6 *†	62.0 ± 4.8 *†
Liver wt (g)	9.75 ± 0.31	9.04 ± 0.26	5.802 ± 0.209 *†
Adrenal wt, paired (mg)	47.8 ± 2.9	42.0 ± 2.0	27.7 ± 1.0 *†
Kidney wt, paired (g)	1.81 ± 0.053	1.72 ± 0.042	1.25 ± 0.094 *†
Ovary wt, paired (mg)	100.9 ± 4.8	89.3 ± 3.7	75.6 ± 3.7 *†
Uterus wt w/fluid (mg)	327.6 ± 26.1	344.9 ± 27.0	340.8 ± 36.3
Uterus wt wo/fluid (mg)	305.7 ± 25.5	319.2 ± 22.9	301.5 ± 20.8
Thyroxine (ug/dl)	4.81 ± 0.27	1.97 ± 0.27 *	0.66 ± 0.02 *
TSH (ng/ml)	7.88 ± 0.41	42.36 ± 6.14 *	99.41 ± 5.67 *
Diestrus	6	6	5
Proestrus/Estrus	4	4	6
Not cycling	3	0	0

Age and weights shown as mean ± standard error of the mean. Only unadjusted values shown.
[b] = Significant Test for Linear Trend ($p < 0.05$)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

3. Discussion

This study tested three compounds which had not been tested in the transferability study. The female pubertal protocol identified the compound which affects the HPG axis (atrazine), but the results were clouded by the possibility of interference by general toxicity. The weak aromatase inhibitor affected several estrogen-related endpoints and had a clear effect on the thyroid. The effects on the estrogen-related endpoints were not clear, however, and this was most likely due to the dual mode of action of fenarimol as estrogenic and anti-estrogenic, which was not known at the time of selection of this chemical for the study. Fadrozole, another aromatase inhibitor, has been shown by other investigators to be positive in the female

pubertal assay. The compound with unknown effects but expected to be a weak estrogen (bisphenol A) did not show estrogenicity and instead provided evidence that general toxicity does not necessarily produce positive results in this assay (i.e., it provided evidence that the female pubertal assay is specific to endocrine interaction).

The other three chemicals tested in this study were also tested in the single-dose, transferability study. The results from the steroidogenesis inhibitor (ketoconazole) confirmed that the same dose level could be identified in a second contract laboratory as interacting with the endocrine system, thus confirming transferability and reproducibility of the assay. The attempt to explore sensitivity by examining a lower dose level of this steroidogenesis inhibitor showed that the adrenals clearly responded but that other endocrine endpoints did not. The results with the estrogen receptor agonist, methoxychlor, also confirmed transferability and reproducibility by showing clear effects at lower dose levels than had caused effects in the transferability study. The exploration of sensitivity with this chemical clearly showed that endocrine effects can be detected by the female pubertal assay in the absence of signs of general toxicity. Finally, the results on the thyroid-system-interactive substance, propylthiouracil, confirmed transferability and reproducibility of results when using this assay, and extended those results by showing that the thyroid effects are clearly discernible at a dose level that is not associated with general toxicity.

B. Multi-dose-level study (TherImmune 2)

1. Purpose

The purpose of this study was to examine the sensitivity of the female pubertal assay to a lower dose level of a previously tested estrogen (methoxychlor), and to a weakly thyroid-active compound which had not been tested in the assay before (phenobarbital). Ethynyl estradiol was included as the positive control for the estrogen.

2. Background and hypotheses for the chemicals

a. Methoxychlor

Methoxychlor has been used for nearly 50 years for insect and larval control (Kapoor *et al.* (1970)). The metabolism of methoxychlor yields mono- and bis-hydroxy metabolites (Bulger *et al.* (1978)), which contributes to both the uterotrophic effects noted earlier for the parent compound (TULLNER (1961)) and the observations that methoxychlor *in vivo* reduced the uterine uptake of radiolabeled estradiol (Welch *et al.* (1969)). Treatment of rats with methoxychlor at 5, 50, or 150 mg/kg for the week before and the week after birth resulted in unchanged anogenital distance but accelerated vaginal opening, and at 50 and 150 mg/kg disrupted adult estrous cyclicity. Additional background information on methoxychlor can be found in Section VI.A.2.c.

b. Ethynyl estradiol

Ethynyl estradiol is among the most commonly prescribed drugs in the United States. The two major uses are as components of combination oral contraceptives and hormone replacement therapy.

Although oral doses of ethynyl estradiol (10 ug/kg) induced a significant increase in uterine weight in the prepubertal rat, this dose was ineffective for stimulating a similar response in ovariectomized adult rats (Laws *et al.* (2000a)). The age of vaginal opening was advanced following oral exposure from postnatal days 21-35 to ethynyl estradiol (10 ug/kg). It advanced vaginal opening by 6.0 ± 0.18 days (30.6 days in control vs. 24.6 days in treated groups). In addition, the number of 4-5 day estrous cycles was reduced during a 25-day exposure to this dose. Advanced vaginal opening was also demonstrated by Odum *et al.* (1997) using doses of 2-400 µg/kg/day, and Singh *et al.* (1980) using doses of 5 µg/kg/day for 5 days. Singh *et al.* (1980) also showed an advance in the appearance of cornified vaginal cells.

c. Phenobarbital

Phenobarbital is a commonly prescribed antiepileptic barbiturate whose hepatotoxicity and effects on the thyroid have been established. Endpoints of

reproductive and developmental toxicity have not been as well explored, although some data are available. Female rats exposed to phenobarbital *in utero* (40 mg/kg/day administered to the dam) or immediately following birth (20 mg/kg/day on PND 1-8) resulted in delayed vaginal opening (34.6 ± 1.2 days in the control vs. 37.5 ± 1.2 days in the treated group). There were disruptions in the estrous cycle (only 40% of treated rats displayed normal estrous cyclicity, compared to 91% of control rats) and impaired fertility (50% of treated rats were fertile, while 100% of the control rats were fertile). These effects were observed during critical periods for neuroendocrine development: GD 17-20 and PND 1-8 (Gupta *et al.* (1983)). There is limited information describing perinatal exposure to phenobarbital. However, it should be noted that phenobarbital (e.g., DePaolo (1985)) and pentobarbital (EVERETT *et al.* (1950)) have been used routinely to block the ovulatory surge of luteinizing hormone in the rat in studies examining the hypothalamic control of GnRH/LH secretion. In these studies, the effective dose of pentobarbital is less than that required for surgical anesthesia, but a dose response for phenobarbital is not well characterized.

Based on the above-cited literature, the hypotheses of the current study were:

1. Ethynyl estradiol and methoxychlor administration to juvenile female rats will result in advanced vaginal opening, advanced first estrous and onset of estrous cycles, and/or the emergence of persistent vaginal estrus. The three doses of methoxychlor used in this study were 12.5, 25 and 50 mg/kg. The two doses of ethynyl estradiol were 2.5 and 5 ug/kg/day. This study was done concurrently with the RTI multi-chemical study, so the results of testing methoxychlor at 25 and 50 mg/kg/day in that study were not available at the time doses were set in this study. The dose level of 12.5 mg/kg/day was included in this study to test the sensitivity of this protocol even further than was tested in the multi-chemical study.

2. Phenobarbital treatment will result in delayed vaginal opening, reduced reproductive organ weights, and irregular estrous cyclicity in juvenile female rats. Phenobarbital treatment may also cause increased thyroid weights and follicular cell hyperplasia. The doses of phenobarbital used were 25, 50 and 100 mg/kg/day. Since there was very little information on which to base a choice of dose levels appropriate for

the juvenile/pubertal female rat, dose levels were chosen to match those administered in a parallel study using male pubertal rats. The dose levels there were chosen based on a finding of an effect by O'Connor *et al.* (1999) in adult male rats and the expectation that juvenile/pubertal animals would be at least as sensitive, or more so, to thyroid-active agents than adults.

3. Results

The final report of this work is presented in Appendix 8[3], and a summary table showing the results for all of the endpoints is included as Appendix 9.

a. Bodyweight

Neither terminal body weight nor body weight gain were different from controls for any of the three compounds tested in this study. (See Table 15, Table 16, and Table 17.) However, because ethynyl estradiol (highest dose) and methoxychlor (all three doses) resulted in earlier vaginal opening, the body weights of these females at VO were significantly lower than controls. In contrast, the body weight of the high dose phenobarbital females (who had a significant delay in vaginal opening) was significantly greater than controls on the day of VO. Finally, it should be noted that a drop in body weight beyond the MTD was not observed for any of the compounds (see Table 4). In fact, it appears that MTD might not have been reached. Demonstration of interaction of these compounds with the endocrine system even without challenging with an MTD provides evidence that the pubertal assay is sensitive to low dose levels.

[3] The final report was written by the contractor based partly on analyses of covariance (ANCOVA) with *terminal* body weight, as requested by EPA. The analyses presented in this Integrated Summary Report, however, are based on analyses of convariance with *weaning* body weight. There may be therefore be differences between the conclusions in the contractor's final report and the conclusions presented here. For a discussion of why ANCOVA with body weight at weaning is preferred, see Section X.E. The ANCOVA analyses using body weight at weaning as covariate are attached as Appendix 10.

Table 15. Methoxychlor: Body weight gain and pubertal development

Parameter	Methoxychlor (mg/kg/day)			
	0	12.5	25	50
Age at VO (PND)	31.9 ± 0.322	27.9 ± 0.215 *†	27.0 ± 0.195 *†	26.5 ± 0.13 *†
BW at VO (g)	117.58 ± 2.482	92.41 ± 2.255 *†	87.48 ± 2.214 *†	82.02 ± 1.217 *†
BW PND 22 (g)	58.87 ± 0.755	58.97 ± 1.003	58.32 ± 1.151	57.47 ± 0.998
BW PND 42 (g)	169.83 ± 2.002	165.79 ± 2.330	171.03 ± 3.441	163.54 ± 1.823
BW as % of control		97.62	100.7	96.30
BW gain 22-42 (g)	111.0 ± 1.769	110.8 ± 2.355	113.5 ± 3.080	106.8 ± 1.717

Mean ± SE (n = 15)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

Table 16. Ethynyl estradiol: Body weight gain and pubertal development

Parameter	Ethynyl estradiol (mg/kg/day)		
	0	0.0025	0.005
Age at VO (PND)	31.9 ± 0.322	31.1 ± 0.63	28.4 ± 0.22 *†
BW at VO (g)	117.58 ± 2.482	112.05 ± 3.967	95.20 ± 1.797 *†
BW PND 22 (g)	58.87 ± 0.755	58.17 ± 0.884	58.03 ± 0.999
BW PND 42 (g)	169.83 ± 2.002	169.01 ± 2.808	171.55 ± 3.728
BW as % of control		99.52	100.1
BW gain 22-42 (g)	111.0 ± 1.769	112.7 ± 2.576	106.1 ± 1.500

Mean ± SE (n = 15)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

Table 17. Phenobarbital: Body weight gain and pubertal development

Parameter	Phenobarbital (mg/kg/day)			
	0	25	50	100
Age at VO (PND)	31.9 ± 0.322	33.1 ± 0.42†	32.8 ± 0.5	34.5 ± 0.67 *†
BW at VO (g)	117.58 ± 2.482	122.63 ± 3.311†	120.67 ± 4.548	128.4 ± 5.107†
BW PND 22 (g)	58.87 ± 0.755	56.70 ± 0.796	57.96 ± 1.070	58.14 ± 0..996
BW PND 42 (g)	169.83 ± 2.002	170.39 ± 3.703	173.11 ± 3.357	169.15 ± 3.128
BW as % of contr0l		100.33	101.93	99.6
BW gain 22-42 (g)	111.0 ± 1.769	113.7 ± 3.234	115.1 ± 2.725	111.0 ± 2.786

Mean ± SE (n = 15)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

b. Vaginal opening

As indicated above, mean day and weight at vaginal opening (VO) in control females were PND 31.9 and 117.58 g, respectively. All three test articles affected the age at VO. Ethynyl estradiol advanced the age at VO to PND 28.4 at 0.005 mg/kg/day. Methoxychlor advanced the age at VO in a dose-related fashion to PND 27.9, 27.0 and 26.5 in 12.5, 25, and 50 mg/kg/day females, respectively. The highest dose of phenobarbital, 100 mg/kg/day, delayed VO to PND 34.5. These were the expected results, and they demonstrate that this endpoint is functioning appropriately in the female pubertal assay.

It is important to note that the LOEL for reproductive effects of ethynyl estradiol in the female pubertal was 5.0 ug/kg/day. This LOEL compares favorably with that reported for the multigenerational study with this compound (National Toxicology Program (NTP) (2007), draft), where a dose of 6 ug/kg/day advanced VO in the F_1 generation. This suggests that the pubertal assay was comparable in sensitivity to the NTP study for this compound.

It is also interesting that Marty *et al.* (1999) reported that estradiol-17β advanced VO in a female pubertal study, but that the dose required was much greater (2 and 4 mg/kg, and a NOEL of 100 µg/kg/day) than the dose required for ethynyl estradiol, reflecting the fact that ethynyl estradiol is more readily available through the gut than estradiol-17β. The consistency of the Marty *et al.* (1999) results on estradiol and the results of this study on ethynyl estradiol confirms that the female pubertal assay is transferable and implies that it is reproducible, even though this was not a direct test of reproducibility.

Even the lowest dose of methoxychlor decreased the age at VO in this study. This is in agreement with the results of Laws *et al.* (2000a) discussed above (Section VI.A.2.c).

Finally, as a barbiturate phenobarbital is known to interfere with the ovulatory surge of LH via a central (hypothalamic) mechanism (DePaolo (1985)). However, it was predicted that this compound would be relatively weak in its ability to interfere with VO based on the dose that is typically used to block the surge in the rat (100 mg/kg) and

the fact that it has been reported that tolerance to phenobarbital does develop with continued treatment (Mycek *et al.* (1976)). The fact that phenobarbital did delay VO and interfered with regular cycles is within expectations; and the finding of this effect at 25 mg/kg/day indicates that the female pubertal assay is quite sensitive.

c. Estrous cyclicity

The mean age at first estrus, cycle length, percentage of animals cycling (whether the cycles are regular or not), and percentage displaying "regular" cycles are shown in Table 18. Treatment with ethynyl estradiol and methoxychlor affected age at first estrus, cycle length, and/or the percentage of regular cycles. In addition to delaying VO, phenobarbital (100 mg/kg/day) decreased the percentage of females cycling. Although the contractor indicated that this effect was unusual (i.e., the difference from controls was due to altered vaginal cycling), this effect would be expected with a compound that alters the hypothalamic control of LH secretion as phenobarbital does (see previous section). A closer inspection of the data (Appendix 8, p 572) shows that VO occurred late and the effect on vaginal cytology included extended vaginal diestrus.

The method used to depict the estrous cycles and the summary below (Table 18) provide useful information beyond simply listing the number of days in diestus or estrus (as above for the multichemical study). Also, listing vaginal cycles and "regular" vaginal cycles in separate columns offers a more informed summary of the data (see Appendix 8, pp 555 and 556). This information was not included in the summary of the previous study (multichemical) partly because the quality of the smear data was weak (difficult to discern cycles in most rats) and because of the lack of regular cycles in the control females.

Table 18. Estrous cycle data

Test article	Dose (mg/kg/day)	Age at first estrus (PND)	Cycle length (days)	Cycling (%)	Regular cycles (%)
Vehicle	0	33.4	4.9	100	80
Ethynyl estradiol	0.0025	32.5	5.1	93	67
	0.005	28.7*	5.1	100	20*
Methoxychlor	12.5	30.9*	5.2*	100	80
	25	30.7*	6.2*	100	27*
	50	28.6*	5.9	100	20*
Phenobarbital	25	35.6	4.9	100	80
	50	34.0	5.0	93	80
	100	34.5	5.3	73*	67

* = p< 0.05

d. Organ weights

Body weight and organ weights at necropsy are presented in Table 19, Table 20, and Table 21 for ethynyl estradiol, methoxychlor, and phenobarbital respectively.

In addition to altering age and weight at VO, ethynyl estradiol treatment resulted in an increase in the mean adrenal gland weight at both 2.5 and 5 ug/kg/day.

The effects of methoxychlor on adrenal gland and kidney weights were not dose related and not considered treatment-related. Liver, ovary, and pituitary weights were lower than control in 50 mg/kg/day treated females. Decreased pituitary weight following treatment with methoxychlor has been observed previously (Goldman *et al.* (1986)).

Phenobarbital treatment resulted in dose-related higher liver weights and higher adrenal gland weights at 25, 50 and 100 mg/kg/day. Thyroid weight was higher than control in 100 mg/kg/day females. Due to an oversight, T_4 and TSH measurements were not made in this study.

The extrathyroidal mechanisms by which phenobarbital reduces serum T_4 are increased hepatic T_4 glucuronidation and excretion (McClain (1989);Barter *et al.* (1992)). Consistent with this extrathyroidal mechanism, liver weights were increased in phenobarbital-treated rats.

Table 19. Effect of ethynyl estradiol on ovarian cycles and tissue weights

Parameter	Ethynyl estradiol (mg/kg/day)		
	0	0.0025	0.005
Endocrine status[a]			
Diestrus	7	5	4
Proestrus	1	2	2
Estrus	5	7	5
Metestrus	1	1	4
Not cycling	0	1	0
Terminal body weight (g)	171.10 ± 1.953	169.91 ± 2.541	173.54 ± 3.358
Adrenal glands, paired (g)	0.04136 ± 0.0017	0.0501 ± 0.0021 *†	0.0487 ± 0.0013 *†
Kidneys, paired (g)	1.72900 ± 0.0290	1.71570 ± 0.03510	1.69916 ± 0.0298
Liver (g)	8.66695 ± 0.1469	8.41119 ± 0.1629	8.75587 ± 0.1581
Ovaries, paired (g)	0.09299 ± 0.0041	0.09435 ± 0.0043	0.09869 ± 0.0046
Pituitary (g)	0.00861 ± 0.0005	0.00816 ± 0.0006	0.00934 ± 0.0005
Thyroid (g)	0.01649 ± 0.0010	0.01687 ± 0.0008	0.01679 ± 0.0011
Uterus w/o fluid (g)	0.35533 ± 0.0276	0.38714 ± 0.0236	0.36026 ± 0.0201
Uterus with fluid (g)	0.41943 ± 0.0492	0.43567 ± 0.0372	0.39721 ± 0.0339

Mean ± SE (n = 15)
* = Significantly different ($p < 0.05$) using unadjusted weights
† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning
[a]Number of females in each stage of the estrous cycle at necropsy, as characterized by vaginal cytology

Table 20. Effect of methoxychlor on ovarian cycles and tissue weights

Parameter	Methoxychlor (mg/kg/day)			
	0	12.5	25	50
Endocrine status[a]				
Diestrus	7	5	11	4
Proestrus	1	1	0	4
Estrus	5	5	3	2
Metestrus	1	4	1	5
Not cycling	0	0	0	0
Terminal body weight	171.10 ± 1.953	167.85 ± 2.126	172.53 ± 3.009	165.59 ± 1.761
Adrenal glands, paired (g)	0.04136 ± 0.0017	0.04865 ± 0.0013 *†	0.04627 ± 0.0022	0.04476 ± 0.0023
Kidneys, paired (g)	1.72900 ± 0.0290	1.66866 ± 0.0291	1.66640 ± 0.0330	1.66618 ± 0.0324
Liver (g)	8.66695 ± 0.1469	8.12279 ± 0.2185	8.39611 ± 0.2484	7.89805 ± 0.1680 *†
Ovaries, paired (g)	0.09299 ± 0.0041	0.09529 ± 0.0043	0.08504 ± 0.0040	0.07888 ± 0.0034 *†
Pituitary (g)	0.00861 ± 0.0005	0.00790 ± 0.0005	0.00771 ± 0.0003	0.00692 ± 0.0005 *†
Thyroid (g)	0.01649 ± 0.0010	0.01725 ± 0.0006	0.01851 ± 0.0009	0.0169 ± 0.0009
Uterus w/o fluid (g)	0.35533 ± 0.0276	0.34717 ± 0.0229	0.33103 ± 0.0183	0.33157 ± 0.0240
Uterus with fluid (g)	0.41943 ± 0.0492	0.37781 ± 0.0373	0.35801 ± 0.0230	0.43732 ± 0.0560

Mean ± SE (n = 15)

* = Significantly different ($p < 0.05$) using unadjusted weights

† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

[a]Number of females in each stage of the estrous cycle at necropsy, as characterized by vaginal cytology

Table 21. Effect of phenobarbital on ovarian cycles and tissue weights

Parameter	Phenobarbital (mg/kg/day)			
	0	25	50	100
Endocrine status[a]				
Diestrus	7	4	7	5
Proestrus	1	2	2	2
Estrus	5	6	3	6
Metestrus	1	3	3	1
Not cycling	0	0	1	3
Terminal body weight	171.10 ± 1.953	171.36 ± 3.543	175.37 ± 3.151	171.46 ± 2.753
Adrenal glands, paired (g)	0.04136 ± 0.0017	0.05203 ± .0025 *†	0.05261 ± .0016 *†	0.05182 ± .0020 *†
Kidneys, paired (g)	1.72900 ± 0.0290	1.95411 ± 0.2092 †	1.74191 ± .07392	1.74029 ± 0.0670
Liver (g)	8.66695 ± 0.1469	9.78521 ± 0.3127 *†	10.7576 ± .3728 *†	11.32150± 0.3304 *†
Ovaries, paired (g)	0.09299 ± 0.0041	0.09769 ± 0.0040	0.10069 ± 0.0034	0.09420 ± 0.0043
Pituitary (g)	0.00861 ± 0.0005	0.00889 ± 0.0010	0.00790 ± 0.0005	0.00832 ± 0.0004
Thyroid	0.01649 ± 0.0010	0.01765 ± 0.0010	0.01813 ± 0.001149	0.01999 ± 0.0007 *†
Uterus w/o fluid	0.35533 ± 0.0276	0.35147 ± .01817	0.34745 ± 0.0208	0.35189 ± 0.0230
Uterus with fluid	0.41943 ± 0.0492	0.40701 ± 0.03914	0.39323 ± 0.0373	0.40873 ± 0.0361

Mean ± SE (n = 15)

* = Significantly different ($p < 0.05$) using unadjusted weights

† = Significantly different ($p < 0.05$) using weights adjusted for covariance with body weight at weaning

[a]Number of females in each stage of the estrous cycle at necropsy, as characterized by vaginal cytology

4. Discussion

In juvenile females, phenobarbital treatment (PND 22-42/43) was expected to result in delayed vaginal opening and irregular estrous cycling. Treatment did delay vaginal cycling (100 mg/kg/day) and decreased the number of animals showing normal estrous cycles (100 mg/kg). Treatment also increased liver and adrenal weights at all doses tested, and thyroid weight at 100 mg/kg. The lowest observable effect level (LOEL) for an endocrine effect was 25 mg/kg/day for phenobarbital, based on the changes in adrenal weight and the increased liver weight. It should be noted that the effects of this compound on the serum thyroxine is mediated through an increased clearance of the hormone by the liver enzymes. Thus, both the liver and adrenal effects are consistent with an endocrine effect. In addition, this study detected alterations in VO and ovarian cycles prior to necropsy that are consistent with the central effects of phenobarbital. In fact, the dose of 100 mg/kg used routinely to block the LH surge in the adult female has been reported to become ineffective in this regard if the dosing is continued for as little as three days after initiation because of the induction of the very liver enzymes that metabolize this compound (Mycek *et al.* (1976)). It is unknown whether a similar diminution of effect is present in the developing animal, but the fact that VO and cycling were altered at 100 mg/kg/day, a dose level that is apparently below the MTD for general toxicity, demonstrates that the female pubertal assay remains sensitive. The clear results for this relatively weak thyroid-active agent, and the low dose levels at which effects were seen, indicate that the female pubertal assay is transferable and can be run with significant sensitivity for this type of compound in a contract laboratory.

Methoxychlor and ethynyl estradiol are both estrogen receptor agonists and were expected to result in advanced vaginal opening, advanced first estrus and onset of estrous cycling, and/or persistent vaginal estrus. Treatment with methoxychlor or ethynyl estradiol (0.005 mg/kg/day) advanced vaginal opening and first estrus, and disrupted estrous cycling. Ethynyl estradiol treatment also resulted in increased adrenal weights. Methoxychlor (50 mg/kg/day) treatment resulted in increased liver, ovary and pituitary weights. The lowest observable effect levels (LOEL) for an endocrine effect

were 12.5 mg/kg/day for methoxychlor and 0.005 mg/kg/day for ethynyl estradiol. 12.5 mg/kg/day was the lowest dose tested for methoxychlor. No endocrine effect was observed at 0.0025 mg/kg/day ethynyl estradiol.

In summary, this study demonstrated that the female pubertal assay detected changes in the estrogen-dependent endpoints (VO, vaginal cycling) and thyroid-system-related endpoints at relatively low doses of these compounds. The results on ethynyl estradiol are consistent with the results in the initial study examining transferability of the protocol (Section V.B), so even though the exploration of sensitivity of the protocol using this compound at a lower dose level did not find effects at that level, the study confirmed that the results at the same dose level were repeatable. The results on methoxychlor were consistent with both the initial transferability study (again, see Section V.B) and the multi-chemical study (Section VI.A.2.c), indicating that the results of the female pubertal assay are reproducible in different contract laboratories when the same dose levels are used. This multi-dose study also showed the sensitivity of the female pubertal protocol for this estrogenic compound, producing effects at one-eighth the level originally tested in the transferability study. The results were produced at doses that did not affect body weight, suggesting that in the female pubertal assay, doses can be found that will result in identification of interaction with the endocrine system without interference from effects of estrogens on body weight.

C. Special studies

EPA's Office of Research and Development (ORD) performed several additional studies to investigate the range of applicability of the both the female and the male pubertal assays across different modes of endocrine activity. These studies demonstrated that the pubertal assays will identify not only chemicals that alter steroidogenesis and steroid receptor binding but also compounds that alter the hypothalamic-pituitary control of gonadal function and thyroid function. In addition, ORD also addressed another technical concern that had generated much debate: whether a reduction in body weight gain during the study period has a confounding effect on the endpoints associated with pubertal development. To address this concern, ORD

evaluated the effects of food restriction on the endpoints in the pubertal both the male and female protocol.

1. Hypothalamic-pituitary-gonadal axis studies

Prior to the multi-chemical study, one of the concerns was whether the female pubertal protocol would be able to detect compounds which targeted the hypothalamic regulation of pituitary hormones. To assist with this concern, ORD scientists performed a study using atrazine, a chlorotriazine herbicide, known to alter hypothalamic GnRH regulation of luteinizing hormone (Cooper *et al.* (2000)).

Atrazine, at doses of 50, 100, and 200 mg/kg/day in the female pubertal protocol using Wistar rats, delayed vaginal opening and altered vaginal cyclicity and other reproductive tract measures (Laws *et al.* (2000b)). The NOEL in this study was 25 mg/kg/day. These observations were replicated in an independent laboratory using both the Alderly Park (Wistar) and Sprague-Dawley strains of rats (Ashby *et al.* (2002)), where the NOEL was 30 mg/kg/day for the Alderly Park strain and 10 mg/kg/day for the Sprague-Dawley strain. The similarity of NOELs between the laboratories and the strains confirms both the transferability and the reproducibility of the female pubertal protocol.

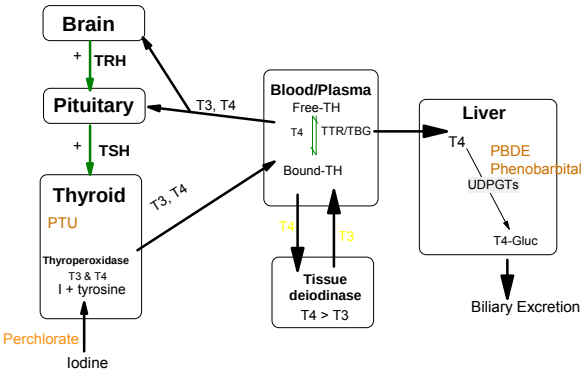
In addition, the primary metabolites of atrazine were evaluated using this protocol (Laws *et al.* (2003)). Although the male pubertal protocol appeared to be sensitive to lower doses of atrazine and the metabolites (Stoker *et al.* (2000);Stoker *et al.* (2002)), the female protocol showed more-robust changes with this chemical (5 day delay in vaginal opening following exposure to 200 mg/kg of atrazine as compared to 3 days delay in preputial separation, the analogous marker of puberty in the male at the same dose level).

Given the pivotal role of the CNS in regulating the onset of puberty, it is not surprising that environmental chemicals that modify the hypothalamic regulation of the pituitary and gonads will also alter the maturation of the gonadal axis. However, this is the first characterization of how such chemicals may be detected with a standardized protocol. It demonstrates that the female pubertal assay is sensitive to HPG-axis modulators.

2. Thyroid axis studies

Chemicals may influence thyroid hormones in one or more ways as outlined in Figure 1. Perhaps the most common mechanism involves increasing the peripheral metabolism of thyroid hormones through induction of hepatic microsomal enzymes as, for example, the polybrominated diphenyl ether DE-71 and phenobarbital do. As a result of this drop in T_4, serum TSH secretion is increased and hypertrophy of the thyroid follows. Other compounds such as propylthiouracil and mancozeb inhibit thyroperoxidase and thus the production of T_4 while perchlorate inhibits iodine uptake. Again, both these compounds will reduce serum T_4 and increase serum TSH. Another potential pathway for impairing thyroid axis function is the disruption of TRH regulation of TSH production. There is evidence that the dithiocarbamate thiram may lead to a decrease in thyroid hormone production through this mechanism. Finally, thyroid hormone economy also can be disrupted by xenobiotics that inhibit the 5'-monodeiodinase, which converts T_4 in peripheral sites (e.g., liver and kidney) to biologically active T_3. (Capen (1997);McClain *et al.* (1989);Pazos-Moura *et al.* (1991);Zoeller *et al.* (2007))

Figure 1. Mechanisms of thyroid hormone disruption



To further evaluate the utility of the female pubertal protocol for detecting environmental chemicals that disrupt thyroid function, additional studies were conducted at the Reproductive Toxicology Division, ORD/EPA. DE-71, a commercial mixture of polybrominated diphenyl ether, induces hypothyroxinemia by increasing the activity of

hepatic microsomal phase II enzyme uridine diphospho-glucuronosyl transferase (UDPGT) that results in an accelerated release of T_4 into the bile (Capen (1997)). Using the female pubertal protocol, DE-71 caused a significant decrease in serum T_4 in Wistar rats following exposure to 30 or 60 mg/kg for 21 days (Stoker *et al.* (2004)). Elevated serum TSH confirmed hypothyroxinemia, and increased liver-to-body weight ratios were highly correlated with a significant induction of UDGPT in these dose groups. Decreased colloid area and an increased follicular cell height were also observed in the 60 mg/kg dose group.

Another chemical tested in the female pubertal protocol was ammonium perchlorate, a sodium iodide symporter inhibitor that decreases circulating concentrations of T_3 and T_4 (Capen (2001);Wolff (1998)). In this study, 125, 250 and 500 mg/kg ammonium perchlorate caused a significant dose-dependent decrease in serum T_3 and T_4 (Laws (2007), personal communication). (See Table 22.) TSH was elevated in the 250 and 500 mg/kg groups. The histological endpoints were most sensitive to treatment as shown by a significant dose dependent increase in follicular height and decrease in colloid area for all doses of perchlorate (62.5 – 500 mg/kg). Finally, a dithiocarbamate pesticide, thiram (tetramethylthiuram disulfide) was tested. Serum T_3, T_4 and TSH were significantly increased in the lowest treatment group (12.5 mg/kg). However, TSH concentrations were 67.8% and 50% of the control for the 50 and 100 mg/kg groups, respectively. While differences in the histological endpoints were minimal among the treatment groups, TSH concentrations were correlated with a significant decrease in follicular height and increase in colloid area for the two highest dose groups (50 -100 mg/kg). These findings present a different hormone pattern from that observed for other dithiocarbamate pesticides such as mancozeb, metiram and zineb that have been reported to alter thyroid function via a common metabolite, ethylenethiourea (ETU), that possibly inhibits thyroid peroxidase (Marinovich *et al.* (1997)). Thiram, however, is not metabolized to ETU (Aldridge *et al.* (1978)), and the thyroid hormone profile observed here suggests a different cellular mechanism.

Table 22. Serum T_3, T_4 and TSH concentrations in female Wistar rats at PND 41 from ammonium perchlorate, DE-71, thiram or atrazine

Chemical	Dose (mg/kg)	Serum T_3 (ng/dl)	Serum T_4 (ug/dl)	Serum TSH (ng/ml)
Ammonium Perchlorate	0	150.1 ± 27.2 (15)	4.90 ± 1.0 (15)	0.82 ± 0.18 (14)
	62.5	147.8 ± 38.5 (15)	4.16 ± 0.90 (15)*	1.31 ± 0.38 (14)
	125	113.7 ± 15.8 (15)*	2.85 ± 0.75 (15)*	1.66 ± 0.51 (15)
	250	115.0 ± 19.5 (15)*	2.32 ± 0.69 (15)*	2.15 ± 1.17 (15)*
	500	108.6 ± 17.6 (14)*	1.72 ± 0.57 (14)*	2.20 ± 1.72 (13)*
Polybrominated Diphenyl ether (DE-71)	0	179.1 ± 55.8 (15)	9.49 ± 2.35 (15)	0.98 ± 0.73 (15)
	3	158.9 ± 43.0 (15)	8.49 ± 2.44 (15)	0.75 ± 0.41 (15)
	30	146.1 ± 49.7 (15)	2.52 ± 1.04 (15)*	1.13 ± 0.74 (15)
	60	154.2 ± 49.7 (15)	1.98 ± 1.49 (15)*	1.44 ± 0.79 (15)
Thiram	0	111.3 ± 25.5 (15)	4.99 ± 1.64 (15)	1.12 ± 0.61 (15)
	12.5	150.0 ± 29.2 (15)*	6.44 ± 1.59 (15)*	1.61 ± 0.44 (15)*
	25	122.7 ± 34.7 (15)	4.68 ± 1.46 (15)	1.13 ± 0.27 (15)
	50	97.4 ± 16.9 (15)	3.86 ± 0.56 (15)	0.76 ± 0.25 (15)*
	100	108.2 ± 13.1 (15)	3.91 ± 0.52 (15)	0.56 ± 0.20 (15)*
Atrazine	0	133 ± 14 (15)	6.02 ± 0.35 (15)	0.57 ± 0.06 (15)
	12.5	144 ± 14 (15)	6.58 ± 0.62 (15)	0.55 ± 0.09 (15)
	25	152 ± 9.0 (15)	6.77 ± 0.35 (15)	0.53 ± 0.04 (15)
	50	129 ± 9.0 (15)	6.05 ± 0.44 (15)	0.62 ± 0.07 (15)
	100	124 ± 10.0 (15)	6.48 ± 0.46 (15)	0.56 ± 0.07 (15)
	200	135 ± 9.0 (15)	6.88 ± 0.28 (15)	0.58 ± 0.06 (15)

T_3 (tri-iodothyronine); T_4 (thyroxine)
Values shown as mean ± SD (n)
Ammonium perchlorate, CAS 20,850-7; Polybrominated diphenyl ether (DE-71), mixture (50-60% penta-BDE, 25% tetra-BDE, 4-8% hexa-BDE); Thiram, CAS 137-26-8
* Significantly different from control, ANOVA, Dunnett Multiple Comparison Test, $p<0.05$

3. Dose selection and body weight issues

The current protocol requires one high dose level at or just below the maximum tolerated dose (MTD), and a lower dose level at half the high dose level. As there is likely to be a minimal amount of information concerning the toxicity of many of the compounds that will be examined in the Tier 1 battery, the basis for the MTD will be based on a decrease in body weight. The MTD for body weight is defined as the dose that produces a 10% reduction in body weight as compared with the appropriate control group (Hodgson (1987)). This approach assumes that a 10% reduction in body weight alone would not alter the endpoints in the pubertal assays. However, one concern with the female pubertal protocol is that many of the endpoints included may be sensitive to alterations in body weight *per se* and thus changes in body weight associated with exposure to the test chemical may confound the interpretation of the data.

There is little doubt that rigorous food restriction regimens resulting in body weight decreases of greater then 50% vs. control will produce moderate to severe reproductive alterations in organ weights, fertility, and reproductive development (Merry *et al.* (1979);Merry *et al.* (1985);Bronson *et al.* (1990);Hamilton *et al.* (1986);Widdowson *et al.* (1960);KENNEDY *et al.* (1963);Perheentupa *et al.* (1995)). However, these studies provide limited insight into how relatively small but statistically significant losses of body weight may influence reproductive and thyroid endpoints. In this regard, several studies have shown that minor reductions in daily food consumption (e.g., approximately 10% of pair-fed controls) and associated weight loss of approximately 10% or less do not alter the endpoints included in the proposed pubertal assays (Engelbregt *et al.* (2001);Chapin *et al.* (1993);Aguilar *et al.* (1984);Ronnekleiv *et al.* (1978)). These studies support the practice of using body weight as a part of the dose-setting process. However, as this assumption had not been fully examined for animals of the specific sex and age used in the pubertal protocols, EPA/ORD conducted a special study to evaluate the effect of food restriction and body weight loss on the parameters measured in the pubertal assays (Laws *et al.* (2007)).

Weanling female rats were provided a daily food supply that was 10, 20, 30 or 40% less than that of controls (determined by actually measuring the control food intake

on each test day) beginning on PND 22 and continuing until PND 42. This regimen of food restriction led to weight loss (when compared to controls) of 2, 4, 12, and 19% respectively at necropsy. Importantly, there was no effect on vaginal opening (even with a 19% decrease in body weight on PND 41-42, see Table 23). Also, the thyroid hormone and TSH concentrations in the serum of the underfed females were not different from controls, even at 19% body weight loss (Table 24). The results also demonstrated that the most critical endpoints in the female assay (with the exception of ovarian weight) were not affected following restrictions in food intake that led up to a 12% reduction in terminal body weight.

Importantly, it is generally accepted that certain organ weights (e.g., liver) are body weight-dependent, and that expression on a relative-to-body-weight basis will correct for body weight decrements (Feron *et al.* (1973)). In the ORD study, a statistically significant decrease was observed in the absolute pituitary, adrenal, liver, and kidney weights in those females in which the terminal body weight was 12% less than controls (Table 25). However, this significance disappeared when the tissues were evaluated on a relative weight basis.

In summary, even though the ORD study did not achieve an exact 10% loss in body weight, the results of this study do support the concept that a 10% reduction in body weight is a reasonable basis for setting the maximum dose for the female pubertal assay. Caution should be used when a reduction in body weight of greater than 10% is observed.

Another possible caveat in selecting a dose based upon the MTD is that estrogen is a known anorexic in rats (see, e.g., Reynolds *et al.* (1974)). Although the effect of estrogen on food intake in prepubertal females has not been clearly defined, it is reasonable to assume that a reduction in body weight would occur if a test chemical is estrogenic. If such a reduction in body weight is mistaken for systemic toxicity and the dosage selected for the protocol reduced accordingly, then selecting a dose based upon the standard MTD criteria may produce false negatives in the overall assay results: a dose would be selected that is too low to demonstrate an endocrine-mediated effect in the protocol. This concern has been mitigated by the results from ethynyl estradiol and

methoxychlor in the multi-dose-level study (Section VI.B.3) in which these estrogenic substances caused clear effects in the absence of changes in terminal body weight. It should also be noted that changes in age at vaginal opening will also help to distinguish estrogenicity from effects due to feed restriction: an advance in the age at VO indicates estrogenicity, while this marker is unchanged or perhaps slightly delayed from feed restriction.

Table 23. Body weights and age at VO following feed restriction

	Level of dietary restriction				
Parameter	Control	10 %	20%	30%	40%
Body weight loss relative to controls	-	2%	4%	12%	19%
Body weight PND 22 (g)	53.5 ± 1.8	56.8 ± 1.9	57.4 ± 1.5	54.5 ± 1.9	55.7 ± 1.7
Body weight PND 33 (g)	116.7 ± 2.5	117.2 ± 1.9	115.0 ± 2.0	107.7 ± 2.1*	99.9 ± 2.4 *
Body weight at vaginal opening (g)	115.8 ± 2.9	117.8 ± 4.5	112.5 ± 3.4	112.4 ± 3.2	105.3 ± 3.9
Body weight PND 41 (g)	152.3 ± 3.2	149.6 ± 2.4	146.0 ± 2.4	133.9 ± 2.7*	123.6 ± 2.9 *
Body weight at necropsy (PND 41-42) (g)	154.8 ± 3.4	151.6 ± 2.6	147.7 ± 2.6	136.0 ± 3.3*	125.6 ± 3.2 *
Age (PND) at vaginal opening	33.0 ± 0.55	33.3 ± 0.67	32.6 ± 0.73	34.0 ± 0.97	34.5 ± 0.97

Mean ± SEM (n = 13)
* Significantly different from the control ($p < 0.05$)

Table 24. Serum hormone concentrations and reproductive tissue weights in females on PND 42

	Level of dietary restriction				
Parameter	Control	10 %	20%	30%	40%
Body weight loss relative to controls	-	2%	4%	12%	19%
T_4 (ug/dl)	4.577 ± 0.228	5.197 ± 0.393	4.794 ± 0.313	4.184 ± 0.331	4.132 ± 0.215
T_3 (ng/ml)	108.03 ± 5.902	106.03 ± 6.564	102.554 ± 4.34	93.48 ± 3.993	93.11 ± 3.130
TSH (ng/ml)	1.022 ± 0.078	1.114 ± 0.122	0.855 ± 0.072	0.843 ± 0.056	0.772 ± 0.07
Ovary (g)	0.060 ± 0.003	0.056 ± 0.004	0.058 ± 0.002	0.047 ± 0.004*	0.041 ± 0.004*
Uterus + fluid (g)	0.321 ± 0.052	0.288 ± 0.042	0.261 ± 0.023	0.277 ± 0.053	0.218 ± 0.043
Uterus - fluid (g)	0.251 ± 0.020	0.243 ± 0.020	0.236 ± 0.015	0.224 ± 0.031	0.180 ± 0.027

Mean ± SEM (n=13)
* Significantly different from the control ($p < 0.05$)

Table 25. Tissue weights in female rats at necropsy following feed restriction

Level of dietary restriction	Necropsy BWT	Pituitary (mg)	Adrenal (mg)	Liver (g)	Kidney (g)
Control	154.8 ± 3.4	8.43 ± 0.24	44.8 ± 1.4	6.74 ± 0.17	1.49 ± 0.03
10%	151.6 ± 2.6	8.19 ± 0.34	43.7 ± 1.9	6.69 ± 0.09	1.45 ± 0.03
20%	147.4 ± 2.6	8.21 ± 0.36	41.6 ± 2.1	5.84 ± 0.46	1.38 ± 0.03
30%	136.0 ± 3.3*	7.31 ± 0.22 *	38.7 ± 1.4 *	5.61 ± 0.22 *	1.29 ± 0.04 *
40%	123.6 ± 2.9*	6.61 ± 0.26 *	37.2 ± 1.2 *	4.75 ± 0.15 *[r]	1.18 ± 0.03 *

Mean ± SEM (n=13)
* Significantly different from the control ($p < 0.05$).
[r] Significant treatment effect (decrease) using relative tissue weight at necropsy ($p < 0.05$).

VII. Establishing performance criteria

Prior to the data analysis of the interlaboratory validation study (described below), performance criteria were established for the endpoints in the assay. These criteria provide a reference for the determining the quality of the data submitted by the participating laboratories and a means to evaluate the variability and efficacy of each endpoint.

In most contexts, "performance criteria" refer to standards of accuracy and/or precision for a positive control chemical, often a weak one, run simultaneously with the test chemical. Such criteria assure that the performing lab can detect at least a minimum signal. In the case of an apical *in vivo* assay such as the female pubertal assay, however, such an approach to performance criteria is impractical. It is not likely that any one chemical will initiate a response in all the endpoints evaluated (e.g., estrogens will not necessarily alter thyroid function and thyroid toxicants will not necessarily alter ovarian function). Thus, testing only one mode may be insufficient to prove sensitivity to another mode. In addition, one chemical may not stimulate all of the endpoints within a single mode of action. The result is that several positive chemicals would have to be tested each time a test chemical (or set of test chemicals) is run. Considering the animal welfare concerns and other expenses associated with an *in vivo* assay of the size and duration of the female pubertal assay, it was deemed inappropriate to require such multiple weak positive controls.

As an alternative, therefore, accuracy and precision performance criteria for vehicle control animals are being used for the pubertal assay. The precision criteria are particularly important inasmuch as they help to ensure that differences between treated groups and controls can be discerned. Criteria on simultaneous controls were also considered more relevant than a periodic, non-concurrent certification program using weak positive controls.

For these reasons, we established performance criteria using historical control data. Before being considered capable of producing high quality data, the performing laboratory must demonstrate that the personnel can conduct all technical aspects of the protocol and provide control data that meet acceptable standards consistent with their

own historical control data bases, as well as data published by other laboratories. The means and CVs for all endpoints should fall within the range specified by the performance criteria. Providing such information allows an independent reviewer to determine whether or not the study was conducted correctly and a basis for comparison of results across laboratories.

As a significant number of studies using the female pubertal protocol were completed prior to the time that the multi-laboratory comparison study was completed, it was possible to use these data to establish performance criteria for most endpoints. Data from the previous EPA-sponsored contract studies, data from peer-reviewed published experiments conducted by other government agencies, commercial and academic laboratories, and data from unpublished studies conducted within the National Health and Environmental Health Effects Laboratory (EPA/ORD) -- the developer of the assay and the laboratory most experienced with the pubertal assay -- were included in this analysis.

A total of 29 studies were examined and approximately half of these studies employed Sprague-Dawley rats, while the other half of the studies employed Wistar rats. For the Sprague-Dawley strain, a performance criterion for the mean value for control animals was established as the interval covered by the mean ± two standard deviations in the historical controls. This is intended to cover approximately 95% of the values likely to be encountered from acceptable laboratories. For the Wistar strain, data were from a single laboratory and while there was an extensive amount of data, it was not appropriate to set performance criteria based on means and standard deviations obtained in this single laboratory.

To create a performance criterion for the variability associated with the measurement of each endpoint, the coefficient of variation (CV) was also determined for each study, and the mean and standard deviation of the CVs for each endpoint were calculated. For this measure, the data from both strains were included because this measure reflects the ability of the individual laboratory (prosector) to perform each measure and is thus independent of the strain. The maximum acceptable value (i.e., performance criterion) for the CVs for each endpoint was set at the mean plus 1½

standard deviation. The minimum acceptable value was set at zero rather than the mean minus 1 standard deviation, since less variability is always desirable. This criterion, too, is intended to cover approximately 95% of the values likely to be encountered from acceptable laboratories.

Table 26 shows the mean value and the magnitude of two standard deviations for each endpoint for the historical controls. The performance criterion for the mean value of the endpoint when the assay is run in a test laboratory is given as the "acceptable range". The coefficient of variation and the magnitude of 1.5 standard deviations of the coefficient of variation in the historical controls are also shown in the table. The performance criterion for the coefficient of variation for an endpoint is given as the maximum value of the acceptable range.

Table 26. 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.

Also, uterine wet weights were so variable in the cycling female that performance criteria would have been difficult to establish. This endpoint, for which data were collected in the transferability, multi-chemical, multi-dose-level, and interlaboratory comparison studies, has been deleted from the protocol.

VIII. Interlaboratory study to examine reproducibility of the female pubertal protocol

A. Purpose

The main purpose of the interlaboratory comparison study was to evaluate the reproducibility of a chemical's effects on the endpoints included in the female pubertal protocol. Specifically, this exercise examined the ability of three contract laboratories to produce similar results when testing the same chemicals at the same two dose levels using the written female pubertal assay protocol. Although the reproducibility of each endpoint was of interest, the main comparison of concern was whether the weight of evidence of the effects leads to the same conclusion from each laboratory concerning interaction with the estrogen and/or thyroid systems. For the weight of evidence, EPA is not requiring that the assay consistently display a pattern of endpoint responses diagnostic for a particular mode or mechanism of action, but only that thyroid-associated responses not be used to claim consistency with sex-steroid-associated responses or vice versa. The final decision about whether there is the potential for interaction of a test chemical with the endocrine system (the goal of Tier 1 screening) is likely to be based on results from a battery of assays, not on any assay in isolation.

Another purpose was to establish whether laboratories with no previous experience with this protocol could reasonably be expected to meet performance criteria based on historical data generated, for the most part, in experienced labs. Based on

this information, another component of this exercise was to determine if the existing protocol required further optimization.

B. Dose-setting for the interlaboratory study

The EDMVS raised significant issues concerning dose selection while reviewing EPA's plans for the interlaboratory validation study. Some members of EDMVS suggested that it is necessary to see how closely different labs can determine appropriate dose levels at which to test an unknown chemical; other members suggested that the focus of a validation study should be on the capabilities of the assay itself, not on the difficulties of dose-setting. EPA decided that in the validation study, dose levels would be the same across laboratories. The reasoning was that it matters less whether laboratories come up with similar high-dose levels in an interlaboratory comparison study than whether the high dose run for a test chemical is shown in the assay itself to be the maximum tolerated dose.

C. Chemicals tested

The chemicals which were tested were the following:

Methoxychlor at dose levels of 12.5 and 50 mg/kg/day

DE-71 at dose levels of 30 and 60 mg/kg/day

2-Chloronitrobenzene (2-CNB) at doses of 25 and 100 mg/kg/day

Due to the expense of conducting large interlaboratory validation studies in a relatively long *in vivo* assay, no attempt was made in this interlaboratory validation to test all of the various modes of action that might be detectable by the female pubertal assay, nor to establish quantitatively the limits of detection that are reproducible across laboratories for each mode. Instead, in accordance with the strategy announced soon after the EDSTAC report, this interlaboratory study focused only on demonstrating that the assay produces similar results across laboratories for a limited number of chemicals.

Methoxychlor was chosen to demonstrate reproducibility for estrogens. DE-71 was chosen to test the thyroid-related endpoints, and 2-CNB was chosen in an attempt to challenge the assay with a toxic but endocrinologically inactive chemical (see below).

Methoxychlor had been tested in all of the previous "pre-validation" studies, once at the 12.5 mg/kg/day dose level and twice at the 50 mg/kg/day dose level. DE-71 had been tested by Stoker *et al.* (2004) in the female pubertal protocol at both of the dose levels used in this interlaboratory comparison study. 2-CNB had not been tested in this assay previously. Both DE-71 and 2-CNB were also tested in the male pubertal protocol at the same dose levels in the same laboratories in a parallel interlaboratory comparison study for that assay.

The choice of 2-CNB deserves particular explanation. Several members of the EDMVS had raised concerns that the specificity of the female pubertal assay had not been tested -- that is, that no compound had been shown to be negative in the assay at the MTD. Without such proof of specificity, there was concern that the pubertal assay might respond to endocrine responses secondary to other toxicities in addition to direct interaction with the endocrine system. The difficulty in testing specificity, however, is that no chemicals have been tested and shown to be negative for the endpoints used in the pubertal assay -- that is, there is no standard against which the specificity of the pubertal assay could be evaluated.

In a good-faith effort to identify a chemical which caused toxicity but no endocrine effects, EPA searched the literature for reproductive and developmental toxicity studies in rats, reasoning that if other toxicities but no reproductive or developmental effects were seen in such a study, the chemical might not be interacting with the endocrine system. Few candidate chemicals were identified. In most cases, chemicals had been tested in reproductive toxicity studies because related chemicals were known to have such toxicity and not surprisingly, the test chemicals tested positive as well.

2-CNB appeared to be the best, though not ideal, candidate. It caused methemoglobinemia but only decreased spermatogenesis among all of the reproductive system endpoints evaluated in the National Toxicology Program's 13-week rat reproductive toxicity study (Bucher (1993)). This study was an inhalation study. No relevant studies in rats by the oral route were identified. In discussions with the EDMVAC after the interlaboratory validation study of the male pubertal assay was complete, one member noted that the similarity of 2-CNB to dinitrobenzene, a known

testicular toxicant, was troubling and should have suggested that 2-CNB would not be an appropriate compound to test. However, it should be noted that the isomer of dinitrobenzene that is toxic to the testis is the 1,3- isomer, and that the 1,2- isomer appears to be inactive (Blackburn *et al.* (1988)).

D. Results

This interlaboratory comparison study was conducted in three laboratories: Charles River/Argus, Huntingdon and WIL Labs. The summary reports from each of these laboratories are attached as Appendix 13, Appendix 14, and Appendix 15 respectively. Appendix 16 presents the statistical analysis comparing the results from the three laboratories, and Appendix 17 provides a detailed table of the results for each endpoint. A summary table is provided as Appendix 18.

1. Ability to meet the performance criteria

The mean value and coefficient of variation for each endpoint for each of the three labs in the interlaboratory study were examined to determine if they met the performance criteria. (See Table 27 and Table 28.)[4] The mean CV for each endpoint by laboratory is shown in Table 29 to allow a more detailed comparison of CVs across laboratories. Also see Figure 2 which displays graphs of historical Wistar and SD data for the mean and coefficient of variation for each endpoint and compares the three contract studies in the interlaboratory study to the historical controls. These data can also be referred to for the section on strain differences to demonstrate the similarities between these two strains of rats for each endpoint. The interlaboratory study used Sprague-Dawley rats in all three laboratories.

[4] Huntingdon conducted the study in two blocks and thus had two control groups, each of which were compared to the performance criteria. 2-CNB and DE-71 were tested with the first set of control animals; methoxychlor was tested with the second set of control animals.

Table 27. Number of endpoints within each lab that met performance criteria for controls for coefficients of variation in the interlaboratory comparison study

	Argus	Huntingdon 1	Huntingdon 2	WIL
Number of endpoints within the acceptable range for CV (0 to [Mean + 1.5 SD of historical controls)	7/11	9/11	9/11	8/11
Endpoints with high CVs	Final BW Liver Kidneys Ovaries -	Final BW Liver - - -	- - - Ovaries Age at VO	- Liver Kidneys - Age at VO

Table 28. Ability of the laboratories to meet the performance criteria in the interlaboratory study for each endpoint

	Argus		Huntingdon 1		Huntingdon 2		WIL	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Age at VO	yes	yes	yes	yes	yes	no	yes	no
Weight at VO	yes	yes	yes	yes	yes	yes	yes	yes
Final body weight	yes	no	yes	no	yes	yes	yes	yes
Liver	yes	no	yes	no	yes	yes	yes	no
Kidneys	yes	no	yes	yes	yes	yes	yes	no
Uterus, blotted	yes	yes	yes	yes	yes	yes	yes	yes
Ovaries	yes	no	yes	yes	yes	no	yes	yes
Pituitary	no	yes	yes	yes	yes	yes	yes	yes
Thyroid	yes	yes	yes	yes	no	yes	yes	yes
Adrenals	yes	yes	yes	yes	yes	yes	yes	yes
T4	no	yes	yes	yes	yes	yes	yes	yes

Mean = the average +/- 2 times the standard deviation of the historical means.
CV = average coefficient of variation + 1.5 times the standard deviation of the CV from historical data, down to zero.
Yes = the lab met the performance requirement for that endpoint
No = the lab did not meet the performance criterion

Table 29. Coefficients of variation for each endpoint in control rats (%)

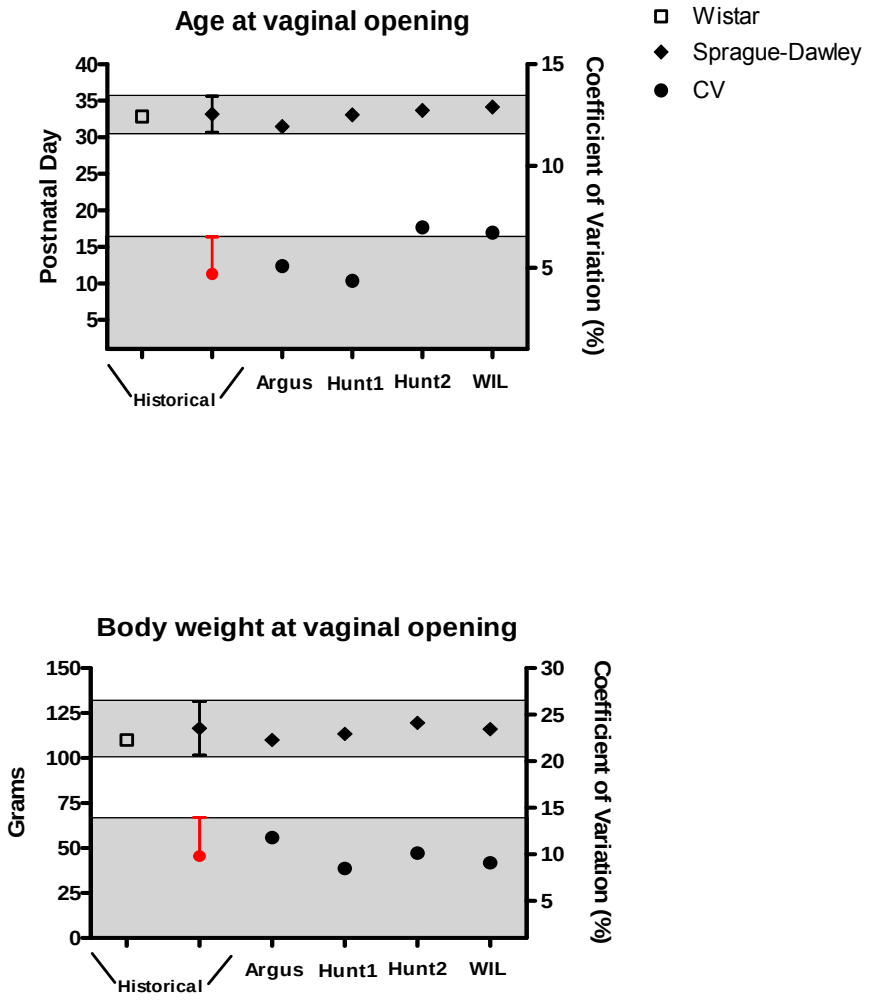
	Argus	Huntingdon 1	Huntingdon 2	WIL
Age at VO	5	4	7	7
Weight at VO	12	8	10	9
Final body weight	10	10	8	8
Liver	16	15	8	15
Kidneys	20	10	8	11
Uterus, blotted	29	21	24	32
Ovaries	26	17	41	22
Pituitary	18	16	19	16
Thyroid	22	11	31	20
Adrenals	22	22	19	9
TSH	33	28	24	26
T4	19	25	13	13

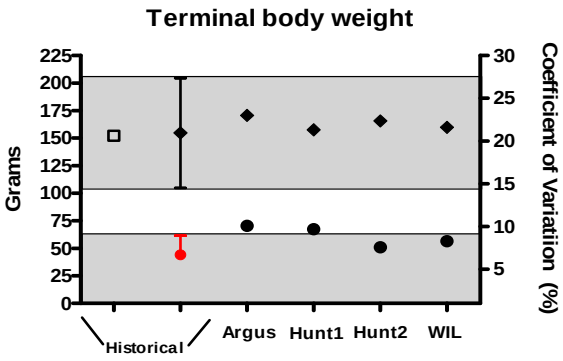
The diagrams in Figure 2 below show how well the individual laboratories met the performance criteria. They plot the historical means for each endpoint, separately for Wistar rats and Sprague-Dawley rats. The range spanned by 2 standard deviations above and below the mean is shown for Sprague-Dawley rats. Similar bars are not shown for Wistars because data were available from only one laboratory. Although there was an extensive amount of data, the range covered by two standard deviations would reflect only intra-laboratory variation and would not be comparable to the range for the Sprague-Dawley rats. To the right of the historical control data, the mean for each lab in this interlaboratory validation study is plotted. Since the rats in this study were Sprague-Dawleys, the acceptable range based on Sprague-Dawley historical controls has been shaded in.

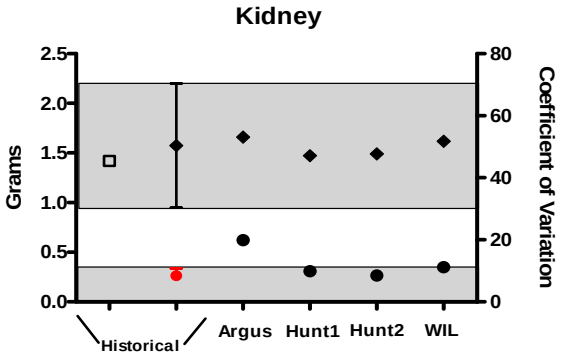
On the same graphs, the coefficients of variation for each endpoint have also been plotted. As explained in Section VII, data from Wistars and Sprague-Dawleys were combined when setting the CV performance criteria so there is only one value for historical controls. The CVs for each laboratory for that endpoint are also plotted. The acceptable range, based on the mean CV + 1.5 standard deviations, is also shown. The acceptable range always reaches to zero since less variability than the historical controls showed is acceptable.

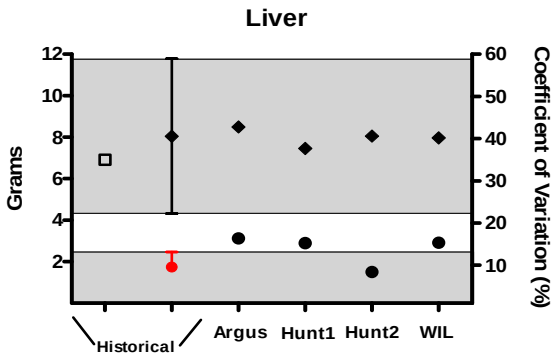
Note that the CVs (i.e., the bottom set of points) are scaled to the right axis, not the left.

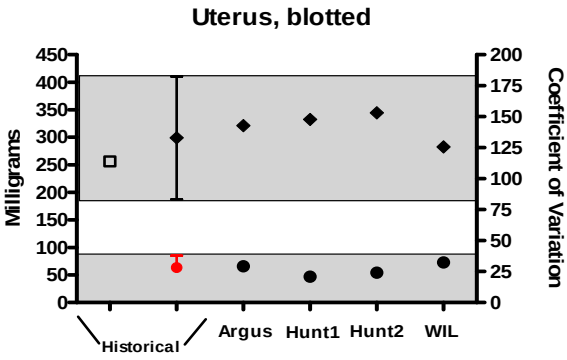
Figure 2. Comparison of the performance of the laboratories in the interlaboratory validation study to the performance criteria

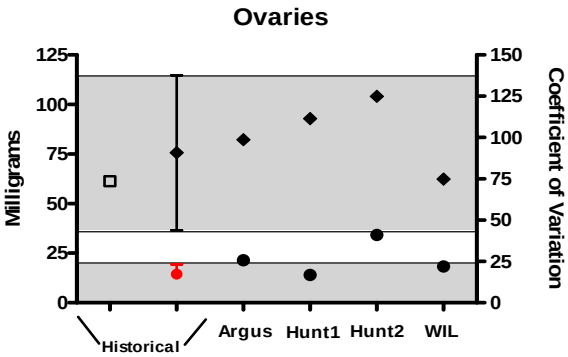


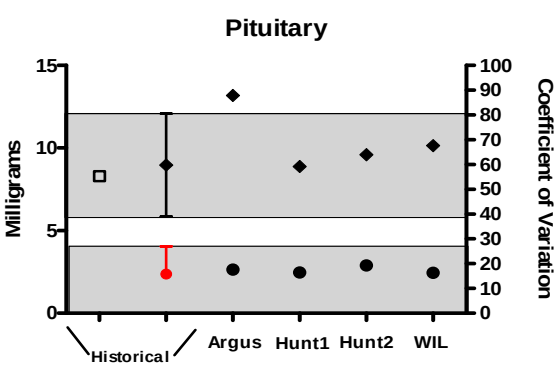

Terminal body weight
Grams
Coefficient of Variation (%)
225
200
175
150
125
100
75
50
25
0
30
25
20
15
10
5
Historical
Argus
Hunt1
Hunt2
WIL

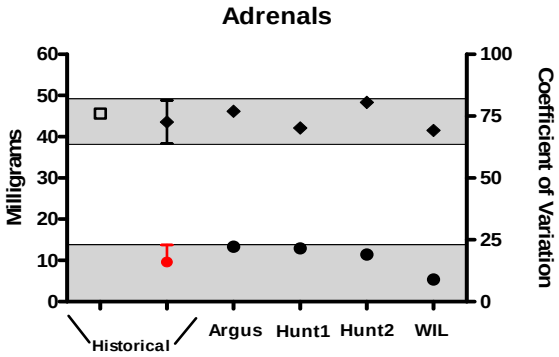

Kidney
Grams
Coefficient of Variation
2.5
2.0
1.5
1.0
0.5
0.0
80
60
40
20
0
Historical
Argus
Hunt1
Hunt2
WIL

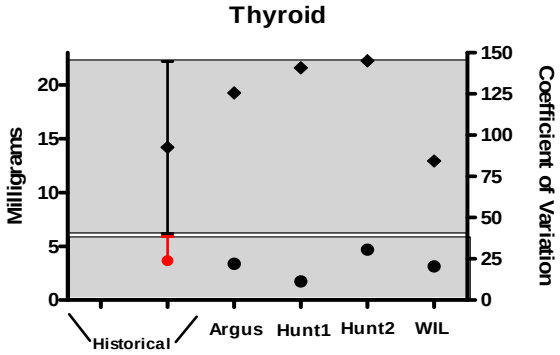

Liver
Grams
Coefficient of Variation (%)
12
10
8
6
4
2
60
50
40
30
20
10
Historical
Argus
Hunt1
Hunt2
WIL

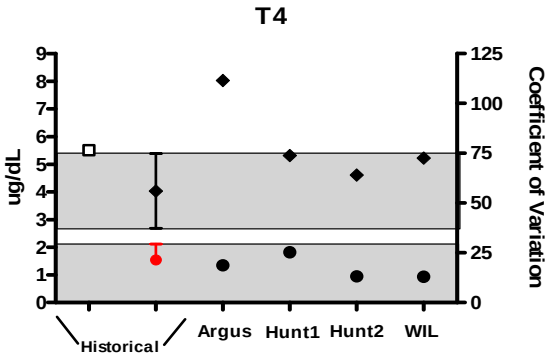

Uterus, blotted
Milligrams
Coefficient of Variation
450
400
350
300
250
200
150
100
50
0
200
175
150
125
100
75
50
25
0
Historical
Argus
Hunt1
Hunt2
WIL


Ovaries
Milligrams
Coefficient of Variation
125
100
75
50
25
0
150
125
100
75
50
25
0
Historical
Argus
Hunt1
Hunt2
WIL


Pituitary
Milligrams
Coefficient of Variation
15
10
5
0
100
90
80
70
60
50
40
30
20
10
0
Historical
Argus
Hunt1
Hunt2
WIL


Adrenals
Milligrams
Coefficient of Variation
60
50
40
30
20
10
0
100
75
50
25
0
Historical
Argus
Hunt1
Hunt2
WIL


Thyroid
Milligrams
Coefficient of Variation
20
15
10
5
0
150
125
100
75
50
25
0
Historical
Argus
Hunt1
Hunt2
WIL


T4
ug/dL
Coefficient of Variation
9
8
7
6
5
4
3
2
1
0
125
100
75
50
25
0
Historical
Argus
Hunt1
Hunt2
WIL

In the discussions of each specific chemical that follow, references to "Laboratory 1", "Laboratory 2", and "Laboratory 3" are to Argus, WIL, and Huntingdon, respectively.

2. Methoxychlor

Methoxychlor is a known weakly estrogenic compound and as noted above has been evaluated extensively in tests for endocrine disruption, including the female pubertal assay. It was expected that methoxychlor would cause a premature vaginal opening at a lower body weight, a reduction in body weight gain, and a smaller terminal body weight.

Table 30 shows that all three laboratories identified early vaginal opening in the methoxychlor treated rats. VO was not affected at the 12.5 mg/kg dose but all three laboratories reported that the 50 mg/kg/day dose induced VO on PND 27. Also, all three labs found that the body weight in the higher dose group was decreased at VO. Early vaginal opening and decreased body weight at VO are significant indicators of estrogenicity and thus methoxychlor was identified as having interaction with the endocrine system at all three laboratories at the 50 mg/kg/day dose level.

Laboratories 2 and 3 reported that terminal body weight was significantly reduced in the high dose methoxychlor rats, both finding that this group weighed 94.1 percent of the controls. This indicates that the Maximum Tolerated Dose was reached, but that the body weight decrease compared to controls was not so severe as to interfere with endocrine endpoints (see Section VI.C.3). Laboratory 1 did not identify a statistically significant change, but did report essentially the same percentage decrease in body weight in the 50 mg/kg dose.

Thus, all three laboratories did identify a similar pattern of response for this weak estrogen and this response was positive for interaction with the endocrine system at the same dose level. This finding is consistent with the finding from the multi-chemical study (Section VI.A.2.c) and the multi-dose-level study (Section VI.B.3) at the same dose level. All three labs were also consistent in finding no response at the 12.5 mg/kg/day level, and although this is not consistent with the TherImmune 2 (multi-dose-level) study (Section VI.B.3), the positive finding at the 50 mg/kg/day level in the absence of other toxicity is sufficient for screening purposes to correctly identify

methoxychlor as having interaction with the endocrine system. (That is, the purpose of Tier 1 screening is not to establish a NOEL, LOEL, or dose-response relationship, so the positive response in the interlaboratory study at 50 mg/kg/day is sufficient.) This interlaboratory validation study confirmed that the female pubertal assay is transferable and provides reproducible responses for a weak estrogenic compound.

3. DE-71

DE-71 is a polybrominated diphenyl ether that is known to alter thyroid hormone function by inducing the hepatic metabolism of this hormone (Stoker *et al.* (2004)). As a result, there is a compensatory increase in TSH secretion and increased synthesis within the thyroid gland, indicated by increased colloidal area and follicular hypertrophy.

All three laboratories in the interlaboratory validation study identified appropriate changes in liver weight, serum T_4 and TSH concentrations. (See Table 30.) With the exception of a decreased pituitary weight and slight but significant (1.61 day) delay in vaginal opening identified by Laboratory 2, there was no other difference observed following exposure to this compound. DE-71 at the lower dose did not affect any of the reproductive parameters. However, whether the effect found by Laboratory 2 in this exercise is indicative of a potential female high-dose reproductive effect, or is simply the result of chance remains to be determined. Regardless of this small discrepancy, the pattern of response in the three laboratories with this compound was very similar.

All three laboratories found decreased colloid area at the high dose level, and two laboratories also observed increased follicular cell height at that dose level. At the low dose level, one laboratory observed both decreased colloid area and increased follicular cell height, while two laboratories observed no significant difference from controls.

This part of the study confirmed that the female pubertal assay produces reproducible and correct results for a weak thyroid toxicant when tested in contract laboratories. The results were consistent across laboratories at both dose levels for almost all endpoints, and with the previously published study on this substance at the same dose levels.

4. 2-Chloronitrobenzene

2-Chloronitrobenzene is used as an intermediate in the manufacture of dyes, rubber and agricultural chemicals. 2-Chloronitrobenzene was chosen with the expectation that it would be toxic to other systems but negative for interaction with the endocrine system (see Section VIII.C). That is, it was anticipated that this compound would induce changes in some other non-reproductive/non-thyroidal endpoint and not alter the pubertal endpoints of concern in this study. However, the compound had never been tested for endocrine effects *per se* before and thus the true response was unknown when this study was initiated

2-Chloronitrobenzene did not have a clear effect on terminal body weight, nor the rate of gain. 2-Chloronitrobenzene treatment did result in a delay in vaginal opening observed by laboratories 2 & 3 and a non-significant delay by laboratory 1 (3.3 days). There was also an increase in body weight at vaginal opening in all three laboratories; as with age at VO, this increase was significant in laboratories 2 & 3 and not-significant in laboratory 1. Ovarian weight, pituitary weight and uterine weight were also reduced following exposure to this compound at the high dose, with significance generally observed in laboratories 2 and 3 and a similar trend in laboratory 1. Liver weight was observed to be significantly increased at both doses in all three laboratories. This effect on liver weight may indicate that the changes in thyroid hormone concentrations were possibly due to an induction of the p450 enzymes associated with T_4 metabolism. T_4 concentrations were decreased at both doses, but were reported significant only in laboratories 1 and 3 for the high dose and laboratory 3 for low dose of 2-chloronitrobenzene. A general increase in TSH was also noted in all laboratories, but this increase achieved statistical significance only in laboratories 2 and 3. An examination of the coefficients of variation for these data suggests that the high variability associated with the measurement of these hormones may account for this discrepancy in achieving a uniform statistical significance across the three participating laboratories.

The finding of effects in all of the labs at a dose which apparently caused no other toxicity suggests that 2-chloronitrobenzene interacts with the endocrine system.

All three labs identified interaction with both the thyroid system and with estrogen function, although the mechanism is not clear.

In Table 30 below, the results for each endpoint are expressed as percent of the control value, with the exception of the mean age of vaginal opening which is listed as the postnatal day this event occurred in the test females. The significant changes are depicted in bold along with an arrow up or down depending on whether or not the compound increased or decreased the response when compared to controls. The table is based on values adjusted for bodyweight at weaning (except for the hormone levels, which are not adjusted).

Table 30. Comparison of effects observed using the female pubertal protocol following exposure to three test chemicals

	MXC 12.5	MXC 50	DE-71 30	DE-71 60	2-CNB 25	2-CNB 100
Final bodyweight[a]						
Argus	96.1	94.8	97.9	99.6	100.4	99.2
Huntingdon	96.2	**94.1↓**	103.7	103.2	101.3	98.5
WIL	100.5	**94.2↓**	101.2	100.1	**105.4↑**	101.5
Bodyweight gain						
Argus	94.8	**92.0↓**	97.8	100.3	101.2	99.6
Huntingdon	95.1	**91.2↓**	103.0	102.2	99.7	95.8
WIL	99.7	**91.1↓**	102.5	100.4	**107.8↑**	102.6
Vaginal opening, age (postnatal day)						
Argus (31.5)[b]	32.0	**27.6↓**	32.5	32.3	32.1	**34.9↑**
H'don (32.9)	32.4	**27.4↓**	34.0	33.1	33.9	**38.7↑**
WIL (32.3)	34.2	**27.4↓**	33.2	**34.2↑**	33.0	**35.1↑**
Vaginal opening, bodyweight						
Argus	101.0	**76.7↓**	105.2	104.4	105.9	**116.6↑**
Huntingdon	**91.9↓**	**67.9↓**	103.3	99.4	102.4	**120.6↑**
WIL	99.5	**65.9↓**	106.8	**110.0↑**	**108.1↑**	**114.1↑**

	MXC 12.5	MXC 50	DE-71 30	DE-71 60	2-CNB 25	2-CNB 100
Ovaries, weight						
Argus	104.5	93.8	94.8	97.4	110.6	90.7
Huntingdon	97.0	**86.2↓**	91.8	89.1	**83.0↓**	**80.3↓**
WIL	98.0	94.5	98.9	92.6	98.5	**86.2↓**
Pituitary, weight						
Argus	84.3	97.9	91.8	82.5	91.8	91.2
Huntingdon	92.2	**87.9↓**	100.1	93.8	**78.9↓**	**66.3↓**
WIL	95.9	92.2	94.3	**87.6↓**	97.1	**88.0↓**
Uterus weight, wet						
Argus	123.8	103.5	95.1	86.6	98.4	81.9
Huntingdon	112.4	95.2	99.3	102.7	93.4	76.7
WIL	118.0	116.5	97.3	89.8	96.9	90.0
Uterus weight, dry						
Argus	116.9	102.6	106.6	92.4	96.5	**79.2↓**
Huntingdon	110.9	109.7	100.5	100.9	89.3	**75.3↓**
WIL	102.7	94.3	98.5	86.1	100.7	87.6
Liver weight						
Argus	96.2	91.6	**126.6↑**	**143.9↑**	**129.9↑**	**160.8↑**
Huntingdon	**89.7↓**	**88.4↓**	**127.0↑**	**144.6↑**	**123.4↑**	**149.8↑**
WIL	96.0	**88.6↓**	**124.4↑**	**132.6↑**	**130.9↑**	**154.4↑**
Adrenals, weight						
Argus	102.4	**117.1↑**	104.5	107.6	108.5	98.9
Huntingdon	96.2	95.8	95.9	90.3	85.3	**79.8↓**
WIL	102.1	106.2	102.7	94.9	103.2	**90.2↓**
Thyroid weight						
Argus	91.2	102.6	112.5	114.4	103.9	101.5
Huntingdon	99.6	106.9	105.3	113.2	95.4	96.8
WIL	87.9	94.2	101.0	95.0	91.4	88.6
Kidneys, weight						
Argus	96.3	96.8	103.5	106.2	109.9	110.8
Huntingdon	**94.5↓**	**91.9↓**	98.4	99.9	96.5	99.1
WIL	97.2	93.3	102.0	99.0	104.9	104.4

	MXC 12.5	MXC 50	DE-71 30	DE-71 60	2-CNB 25	2-CNB 100
T_4						
Argus	98.4	97.0	**58.6↓**	**52.0↓**	90.1	**86.1↓**
Huntingdon	93.3	94.1	**29.2↓**	**13.6↓**	**77.8↓**	**72.2↓**
WIL	96.6	108.7	**26.0↓**	**17.6↓**	98.2	93.0
TSH						
Argus	119.9	97.9	**147.4↑**	**139.2↑**	126.6	111.1
Huntingdon	121.8	115.7	**134.3↑**	**146.2↑**	106.0	**125.3↑**
WIL	121.6	110.2	**140.8↑**	**174.2↑**	114.6	**128.9↑**

[a] All data, with the exception of the age at vaginal opening, is reported as percent change from control (i.e., treated value divided by control value) so that the direction and magnitude of change can be evaluated in one table. Raw data for this table is presented in Appendix 16.

[b] The age of vaginal opening in control animals for the laboratory is shown in parentheses. Due to the statistical adjustments needed to put all laboratories' data on a comparable basis, there are two sets of control numbers for age at VO. The appropriate control figures to use for comparison to methoxychlor are 31.5, 33.6, and 34.5 for Argus, Huntingdon, and WIL respectively.

E. Conclusions from the interlaboratory comparison study

The current study demonstrates that the female pubertal protocol is transferable and reproducible in contract laboratories. For both the estrogenic and the thyroid-system-interactive chemical, all three laboratories' results showed interaction with the endocrine system. The primary indicators of estrogenicity, age and weight at vaginal opening, decreased markedly in all three laboratories at the same dose level of methoxychlor, the weak estrogen in the absence of other signs of toxicity. Similarly, the primary indicators of interaction with the thyroid system, T_4 and TSH levels, responded markedly in the appropriate manner for the weak thyroid-active agent. For the chemical which had never been tested before, results were consistent across all three laboratories.

Each laboratory had CVs for certain endpoints which were not within the range specified by the performance criteria. In one laboratory, and in one set of controls from another laboratory (out of two sets of data), the criterion for age at vaginal opening was not met but the labs were still able to detect the appropriate change in this endpoint from the estrogen methoxychlor, emphasizing the robustness of this endpoint.

IX. Data interpretation

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 31. 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 31. 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.

X. Additional issues concerning sensitivity, specificity, and reproducibility

This section discusses additional issues that may be of concern in the context of the validation of the female pubertal assay for use in a screening program.

A. Phytoestrogens in feed

One concern that has been raised is that the soy and alfalfa content of most major rat feeds may contain sufficient phytoestrogens to interfere with the sensitivity of endocrine assays (Thigpen *et al.* (2004)). There are reports that some laboratory rodent dietary formulations contain levels of phytoestrogens that are sufficient to induce alterations in uterine weight and histology (Boettger-Tong *et al.* (1998)). As a result,

there has been debate about the potential influence of the diet on the endpoints measured in the pubertal assays.

Relevant to this discussion is the report on the uterotrophic assay in which the phytoestrogen content in the diet of the participating laboratories was analyzed and the impact of the diet on uterine weight examined by an international group of experts for the Organisation of Economic Cooperation and Development (OECD (2006)). In this analysis, no relationship between phytoestrogen content and uterine weight was observed for the adult female but a "suggestive" effect of high concentrations of phytoestrogens on the uterine weight of the immature animal was discussed. The suggestion that the phytoestrogen content may influence uterine weight in the developing animal (such as those used in the female pubertal protocol) was apparently the result of one laboratory's data (laboratory 20), and there was a question about the quality of the data from that laboratory. The following is the relevant excerpt from the report:

> "An analysis of the phytoestrogen contents of the laboratory diets revealed significant levels in many diets. A review of food consumption indicates that this would lead to different dietary intakes on an approximate ratio for OVX [ovariectomized] adult rats:immature rats:OVX adult mice of 1:2:4. An examination of the vehicle control weights and the responses to the weak agonists in different laboratories was made relative to an estimated dietary intake of phytoestrogens. The data indicated that no effect was evident for the adult OVX model. However, the data were suggestive of an effect for the immature rat model when GEN [genistein] intakes would exceed 50 mg/kg/d. This level is consistent with other toxicological studies showing a LOEL in this range as well as the MED values in this study for Protocol A. However, the interpretation that an influence of dietary phytoestrogen interferes with the study results largely relies on the results in Laboratory #20. A close examination of those data reveals that these data are open to question, and any conclusions must be drawn with caution until controlled studies are done with defined diets, defined doses, and sufficient doses of phytoestrogens. However, as a precaution until such data are available, experiments with immature rats or OVX mice should limit the dietary content of phytoestrogens to about 350 µg phytoestrogens/g diet and 175 µg phytoestrogens/g diet, respectively."

As these data are pertinent to the female pubertal protocol, a similar conclusion was made and a limit of approximately 300 ug/g in feed was included in the female

pubertal protocol. However, the same note that "any conclusions must be drawn with caution until controlled studies are done" applies to the pubertal protocol as applies to the uterotrophic assay.

There are few other studies investigating the effect of phytoestrogens in feed on the endpoints measured in the female pubertal assay in rats. Odum *et al.* (2001) investigated several diets and concluded that "choice of diet may influence the timing of sexual development in the rat, and consequently, that when evaluating the potential endocrine toxicity of chemicals, the components of rodent diets used should be known and as far as is possible, controlled." However, they concluded that these differences are more likely to be cause by nutritional differences than by individual dietary components. You *et al.* (2002) concluded that there was accelerated vaginal opening at 300 ppm of genistein (a phytoestrogen) in feed and altered cyclicity in addition to accelerated VO at 800 ppm, but a closer look at the data suggests that the control group had delayed vaginal opening rather than that the genistein group was accelerated. Lower levels were not tested. They concluded that the estrogenic responses to genistein and methoxychlor administered together were apparently accumulative of the effects associated with each compound alone, but also suggested that "the presence of phytoestrogens in the diet of experimental animals needs to be fully considered as part of the experimental conditions, particularly when such experiments are aimed at evaluating endocrine properties of test compounds." Masutomi *et al.* (2003) found no effect of genistein on adrenals, ovaries, or uterus weights at dose levels up through 1000 ppm of genistein, but these weights were measured at postnatal week 11 rather than the pubertal assay's measurements at the end of postnatal week 6. Masutomi *et al.* (2004) reported increased (not decreased) responsiveness of age at vaginal opening to ethynyl estradiol administered through the diet when a standard rodent diet containing soybean-derived phytoestrogens was used, compared to when a soy-free diet was used. There was no effect of diet on response by adrenals, ovaries, or uterus to ethynyl estradiol, again at postnatal week 11.

While Thigpen *et al.* (2003) showed that vaginal opening in mice can be accelerated by phytoestrogens in feed, they did not study whether this interferes with

the detection of interaction of a test substance with the endocrine system. Control animals will be fed the same batch of feed as the test animals in the female pubertal assay, so it is the difference between treated groups and controls that is of interest, not whether the groups have different absolute values than they would have had if raised on a phytoestrogen-reduced diet.

Given this information, it appears that there is still considerable uncertainty, given the complexity of studying the many variables in feeds, about the effect of phytoestrogens in feed on the endpoints in the female pubertal assay. EPA believes it prudent to set a limit on the concentration of phytoestrogens in feed used in the female pubertal assay, but agrees with the OECD that further controlled studies are required to understand the relationship of phytoestrogens in feed to the outcomes of the assay.

B. Rat strain differences

Concerns have been raised both internally and from the EDMVS (one of the Federal Advisory Committees which was consulted during the development of the pubertal assay; see Section I.D) that the strain of rat used may affect the ability to detect a response in endocrine assays.

The limited number of studies employing the female pubertal protocol have not identified obvious strain differences. The studies performed by contractors for prevalidation and validation that are described in this report used Sprague-Dawley rats, while all of the studies performed by EPA employed Wistar rats. The data from control Wistar and Sprague-Dawley rats reveals remarkable agreement for essentially all the parameters measured in the female pubertal protocol, as shown in the performance criteria diagrams (Figure 2). Differences and similarities between results when a test chemical was run at both an EPA lab and a contractor's lab (atrazine, DE-71) have been described earlier in the report, but no systematic differences attributable to strain have been discerned.

Beyond these studies, there is little information to define how robust the protocols are across strains. The response of the female to atrazine is similar in the Wistar (Laws *et al.* (2000b)); Sprague-Dawley (Section VI.A.2.a), and Alpk:ApfSD (Wistar-derived

female) (Ashby *et al.* (2002)) strains. The effects of the endocrine disrupting chemicals on the female Sprague-Dawley versus Long-Evans tested in the transferability study (Section V.B) revealed general agreement between the strains, with the exception that the Sprague-Dawley females had a significant delay in vaginal opening following 240 mg/kg/day of PTU while the Long-Evans females did not. Also, there was a discrepancy in the response of these two strains' thyroid stimulating hormone response to methoxychlor and pimozide in which a significant decrease in TSH (1.77 vs. 0.89 and 1.77 respectively) was noted in the Long-Evans, but not the SD females (1.53 vs. 1.33 and 1.30 respectively). However, the group size in this study was less than half of the standard group size (6 vs. 15), so any conclusions should be viewed with caution.

In summary, these results indicate that the different strains of animals commonly used in toxicity testing do not vary greatly in their response to the pubertal assay. However, there are far too few data to conclude that this is the case with certainty. Recognizing this, and in response to EDMVS' recommendation to write a White Paper on the issue of rat-strain effects on pubertal assay endpoints, EPA prepared such a paper and presented both the White Paper (Appendix 11) and an expert reviewer's comments (Appendix 12) to the EDMVS in August 2003. EPA concluded that although it appears that some strains of rats are differentially sensitive to endocrine effects, it is not possible at this time to determine which strain will be the most susceptible across all (or most) endpoints. Because the pubertal assay has multiple endpoints, it was not possible to choose an optimal strain (if indeed an optimal strain exists). In this, the pubertal assay is no different from any other toxicity test with multiple endpoints. While EDMVS did consider recommending that multiple strains be used, it decided that this would not be feasible for the multi-endpoint pubertal assays.

At this time, there is a preference of Crl:CD®(SD) rat for the female pubertal protocol. While the EPA recognizes there are reasons to believe that this strain might be particularly insensitive to endocrine disruptors (see Appendix 12), the data currently available appear to show that it is no worse (or better) than other strains for screening for endocrine activity using the endpoints in the male pubertal assay. Other considerations therefore form the basis for EPA's preference. This strain of rat is readily

available and there is a reasonable amount of endocrine data available from pubertal studies, thus making it possible to establish performance criteria.

In summary, EPA is aware of the potential for differences between strains and therefore expresses a preference for standardization using the Sprague-Dawley rat. Given the data currently available and the amount of research it is likely to take to determine the best strategy for optimizing the use of rat strains, the current uncertainty about the effect of strain on sensitivity does not disqualify this assay for use in a screening program

C. Specificity

Another concern that has been raised about the female pubertal assay is that no chemicals have been shown to be entirely negative in this assay. The lack of negative reference chemicals (that is, chemicals which have been tested for all the endocrine activities that can be identified by the pubertal female assay and which are known to be negative for all of these activities) made it difficult to test the specificity of this assay. As noted in the section describing the choice of chemicals for the interlaboratory validation study (Section VIII.C), a good-faith effort was made to identify a chemical that was both toxic to other systems but without endocrine effects. Upon testing in this assay, however, the chemical gave positive results. Since at this time it is not known from other assays whether this chemical interacts with the endocrine system, it is not possible to determine whether the pubertal female assay is non-specific or the chemical is indeed interacting with the endocrine system.

It is clear, however, that the female pubertal assay's estrogen-related endpoints do not respond to all stresses. Several chemicals are known thyroid toxicants that have been shown to be positive for the thyroid effects and negative for the endocrine and reproductive effects in the female pubertal assay. One example of this is propylthiouracil at the 2 mg/kg/day dose level (Section VI.A.2.f). This chemical altered thyroid hormones and thyroid weight and caused no effects on any of the reproductive tract weights or puberty onset.

Similarly, the assay's thyroid-related endpoints do not respond to all stresses. The study on atrazine (Section VI.A.2.a) is an example of estrogen-related endpoints being affected while thyroid-related endpoints are not. Thus, there is reason to believe that the assay is specific to interaction with the endocrine system rather than to general stress.

D. Selection of dose levels

One important issue is the selection of appropriate dose levels so that toxicity is avoided. Guidance for determining appropriate dose levels generally focuses on determination of the maximum tolerated dose.

The problem of MTD determination, however, is not unique to the pubertal assay. All assays, for any effect, must show that an adequate challenge has been presented to the system before a negative finding from that challenge can be accepted. MTD determination is not usually considered part of those protocols, nor is the accuracy of MTD determination considered a measure of the validity of those assays.

The EPA recognizes that some investigators may choose to perform special studies to estimate the proper dose levels to use in the pubertal assay. This route may be chosen in some cases because information which is often available for the MTD of adult rats for a 28-day or 90-day exposure may not be applicable to the 20-day exposure in juvenile/pubertal rats on which this assay is based. Such studies are not required, however. The only requirement is that a dose level at or near the MTD be tested before making a claim that the substance does not interact with the endocrine system.

Due to the importance of the MTD determination, EPA has clarified what it will consider evidence of exceeding the MTD for this assay. In addition to clinical observations that indicate stress, terminal body weight loss compared to controls that exceeds approximately 10% and is statistically significantly different from controls may be used as evidence that MTD has been exceeded. Terminal body weight loss that is not statistically different from controls may be an indication that MTD was not reached. In addition, abnormal blood chemistry values at termination (particularly creatinine and

blood urea nitrogen (BUN)) may indicate that MTD was exceeded. Finally, 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.

In some cases the second, lower, dose level required by the pubertal protocol can provide useful information even if the high dose level exceeds the MTD.

E. Adjustment for body weight at weaning

When examining organ weights, researchers often consider the difference in the organ-to-body-weight-ratio (or similarly, analysis of covariance with terminal body weight) as the appropriate indicator of whether treatment affected organ weight. Such analyses separate the effect due to treatment from any effect that may be due simply to size of the animal.

Those analyses, however, are not appropriate when treatment itself may affect body weight. (See Section VI.C.3 for a discussion of endocrine-active compounds and their relationship to body weight.) It would separate the effect on each organ only if that effect exceeded the treatment's effect on body weight.

Since it is not possible to measure what the terminal body weight would have been without treatment, body weight at the last point before treatment is used as a surrogate. In this way, effects of treatment on specific organs are separated from the effect of treatment on overall body weight while still taking into account, as best as possible, the effect of the animal's (untreated) size.

The adjustment of organ weights (and age and weight at VO, but not hormone levels) for covariance with body weight at weaning is an additional measure to take into account untreated size. The procedure described in the protocol for distributing weanlings to test groups randomly based on weight so that all groups are similar in mean and standard deviation at the beginning of treatment also helps to ensure that the group means of treated animals vs. controls for organ weights are not due to differences in untreated size.

XI. Summary of the female pubertal protocol

Validation of multi-endpoint, apical, *in vivo* assays such as the female pubertal assay is limited by the availability of appropriate reference chemicals to test, as well as by time and other resources. Nevertheless, within this context, the female pubertal protocol has proven to be transferable, sensitive, and reproducible over many different modes of action. The following sections on strengths and weaknesses discuss the assay in terms of its potential place in a battery of assays for identifying interaction with the endocrine system.

A. Strengths of the female pubertal protocol

The strengths of the female pubertal protocol in the context of a screening program to identify the ability of a test chemical to interact with the estrogen and/or thyroid system in humans are that it is an *in vivo* assay, it is performed in a mammalian model, it is an apical assay, it has redundant confirmatory endpoints, it involves the pubertal period of development, and it has a well-established base of knowledge of the relationship of the endpoints to endocrine activity. In many cases, the profile of responses across the various endpoints can suggest mechanisms of action that might be operative and this can help focus attention in later studies.

The fact that metabolism is accounted for in an *in vivo* assay provides greater confidence, as compared to an *in vitro* assay. In some known cases such as methoxychlor, it is the metabolites which are the most active agents and *in vitro* systems may not identify the parent compound as having the potential to interact with the endocrine system when taken in by a complete organism. Thus, the use of an *in vivo* system reduces the likelihood of false negative or false positive results. In addition, the integrated nature of the endocrine system in the developing organism and the relationship of the endocrine toxicity to other systemic effects cannot be simulated *in vitro*. Also, the use of a mammalian model usually gives rise to greater confidence that results are relevant to humans than if a phylogenetically more removed model is used.

The fact that this is an apical assay is both a strength and a weakness. The ability to detect in a single assay many different modes of action, as demonstrated

during the validation process, is efficient. For example, the assay provides the opportunity to measure, in one assay, both reproductive and thyroid responses.

Having redundant confirmatory endpoints is helpful in an assay such as this where variability in proficiency in measuring specific endpoints can be a factor in the ability of the assay to detect a response. Performance criteria, particularly for weak positive controls, can help ensure a minimum level of sensitivity, but redundant endpoints provide additional aid.

The pubertal assay focuses on a period of development when the endpoints selected for the assay are particularly sensitive to endocrine modulation. Sensitivity is therefore greater than for other life stages which are feasible to include in a screen for endocrine activity.

One of the strengths particular to the female pubertal protocol as an assay for screening for interaction with the endocrine system lies in the fact that the measures included for the identification of endocrine effects are based on a solid knowledge of how the reproductive and thyroid axis mature in the rat. Thus, the extensive basic literature in this area, reviewed in the Detailed Review Paper (Appendix 2), provides the background for the underlying assumptions in the assay endpoints and assists in the interpretation of results.

The female pubertal protocol consists of several relatively straightforward measures that can be performed in most professional laboratories. This is not to say that training and expertise is not needed to perform the technical aspects of tissue dissection, serum collection, hormone assay, tissue preparation, histology and histopathology. However, all of these procedures are routinely performed in contract laboratories and are an integral part of the current requirements for pesticide registration. Thus, the technical difficulties of this assay should not be a barrier to its implementation under Good Laboratory Practices. In the early stages of prevalidation, laboratories appeared to have difficulty in measuring some of the endpoints in the protocol. For example, the measurements of small tissue weights (i.e., adrenal, pituitary) did eventually require a more explicit description of the tissue dissection procedures in the protocol itself. Also, the variability in hormone data in some of the

early studies was a cause for concern. Many of these early issues did not appear in the later studies once the methods were described more precisely in the protocol and the participating laboratories became aware of the areas of concern.

In many cases, the profile of responses provided by the protocol will make it possible to develop hypotheses concerning the mechanism and mode of action that may be involved in any adverse outcome.

The sensitivity of the female pubertal assay appears to be high. Although it has not been studied thoroughly, the assay appears to provide a good estimate of the NOEL obtained from other studies of the same or longer duration for some chemicals. For example, the NOEL in the female pubertal for the chlorotriazines is below that observed in the adult female treated for the approximately the same period of time (50 vs. 75 mg/kg respectively) (Laws *et al.* (2000b)). Also, as explained in Section VI.B.3.b, it is significant that the LOEL for ethynyl estradiol in the female pubertal assay was comparable to the LOEL for reproductive effects in a recently-completed multigenerational study.

Finally, the development of performance criteria and the incorporation of this information into the final protocol provide important guidance for this work.

B. Weaknesses of the female pubertal protocol

There are limitations to this protocol. Perhaps the most troublesome is the inherent variability of the female organ weights. The female pubertal protocol utilizes ovary-intact animals and thus, once puberty is attained, the fluctuations associated with the normal ovarian cycle could be viewed as problematic as evidenced by the high coefficients of variation. In analyzing the control data for the development of the performance criteria, it was possible to address the question as to whether it would be advantageous to kill all the females on the same day of the estrous cycle rather than the PND 42 as currently required. We did identify a decrease in the variability of the wet and blotted uterine weights (to an average of about 25%), but the reduction did not appear to justify a change in the protocol at this point.

Although identification of mechanism of action is not necessary for identification of interaction with the endocrine system, some observers may feel that the inability of this assay to isolate mechanisms of action is a limitation. For example, although the protocol did detect the adverse effects of the chlorotriazines (Laws *et al.* (2000b)), it would have been difficult to determine whether or not this compound was an anti-estrogen, blocked steroidogenesis or disrupted the central control of puberty on the basis of the female pubertal assay's results alone. For this compound, information available from other studies published prior to the atrazine pubertal study was available and the CNS-hypothalamic effects of this compound were well established. In contrast, the uterotrophic assay is reasonably specific for compounds that interact with the estrogen receptor.

A significant limitation of the EDSP's implementation of the protocol, but not the protocol itself, is the absence of concurrent weak positive controls. As explained in Section VII, requirement of a sufficient number of positive controls to cover all the endpoints and all the potential modes of action for an apical assay appears to be infeasible. This weakness has been mitigated by the inclusion of performance criteria for the controls, but is not eliminated.

C. Conclusion

In summary, EPA believes that the female pubertal protocol has proven to be transferable, sensitive to the kinds of interactions with the endocrine system it claims to detect, and reproducible in independent contract laboratories. It also finds reason to believe that the assay is specific even though this is not testable at this time. While there may be ways to improve the assay in the future, the assay appears to be appropriate for use in a screening program to identify interaction with the endocrine system.

XII. References

1. Aguilar, E., Pinilla, L., Guisado, R., Gonzalez, D., and Lopez, F. (1984). Relation between body weight, growth rate, chronological age and puberty in male and female rats. Rev.Esp.Fisiol. **40**, 83-86.

2. Aldridge, W. N. and Magos, L. 1978. Carbamates, thiocarbamates, dithiocarbamates. V/F1/78/75 EN. Luxembourg, Commission of the European Communities.

3. Andersen, H. R., Bonefeld-Jorgensen, E. C., Nielsen, F., Jarfeldt, K., Jayatissa, M. N., and Vinggaard, A. M. (2006). Estrogenic effects in vitro and in vivo of the fungicide fenarimol. Toxicol.Lett. **163**, 142-152.

4. Andersen, H. R., Vinggaard, A. M., Rasmussen, T. H., Gjermandsen, I. M., and Bonefeld-Jorgensen, E. C. (2002). Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity in vitro. Toxicol.Appl.Pharmacol. **179**, 1-12.

5. Ashby, J. and Tinwell, H. (1998). Uterotrophic activity of bisphenol A in the immature rat. Environ.Health Perspect. **106**, 719-720.

6. Ashby, J., Tinwell, H., Stevens, J., Pastoor, T., and Breckenridge, C. B. (2002). The effects of atrazine on the sexual maturation of female rats. Regul.Toxicol.Pharmacol. **35**, 468-473.

7. Barter, R. A. and Klaassen, C. D. (1992). UDP-glucuronosyltransferase inducers reduce thyroid hormone levels in rats by an extrathyroidal mechanism. Toxicol.Appl.Pharmacol. **113**, 36-42.

8. Blackburn, D. M., Gray, A. J., Lloyd, S. C., Sheard, C. M., and Foster, P. M. (1988). A comparison of the effects of the three isomers of dinitrobenzene on the testis in the rat. Toxicol.Appl.Pharmacol. **92**, 54-64.

9. Boettger-Tong, H., Murthy, L., Chiappetta, C., Kirkland, J. L., Goodwin, B., Adlercreutz, H., Stancel, G. M., and Makela, S. (1998). A case of a laboratory animal feed with high estrogenic activity and its impact on in vivo responses to exogenously administered estrogens. Environ.Health Perspect. **106**, 369-373.

10. Bronson, F. H. and Heideman, P. D. (1990). Short-term hormonal responses to food intake in peripubertal female rats. Am.J.Physiol **259**, R25-R31.

11. Bucher, J. (1993). NTP technical report on the toxicity studies of 2-Chloronitrobenzene (CAS No. 88-73-3) and 4-Chloronitrobenzene (CAS No. 100-

00-5) Administered by Inhalation to F344/N Rats and B6C3F1 Mice. Toxic.Rep.Ser. **33**, 1-F25.

12. Bulger, W. H., Muccitelli, R. M., and Kupfer, D. (1978). Studies on the in vivo and in vitro estrogenic activities of methoxychlor and its metabolites. Role of hepatic mono-oxygenase in methoxychlor activation. Biochem.Pharmacol. **27**, 2417-2423.

13. Capen, C. C. (1997). Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicol.Pathol. **25**, 39-48.

14. Capen,C.C. (2001). Toxic responses of the endocrine system. In Casarett and Doull's Toxicology: the basic science of poisons (C.D.Klaassen, Ed.), pp. 617-640. McGraw-Hill, New York, NY.

15. Chapin, R. E., Gulati, D. K., Barnes, L. H., and Teague, J. L. (1993). The effects of feed restriction on reproductive function in Sprague-Dawley rats. Fundam.Appl.Toxicol. **20**, 23-29.

16. Cooper, R. L., Stoker, T. E., Goldman, J. M., Parrish, M. B., and Tyrey, L. (1996). Effect of atrazine on ovarian function in the rat. Reprod.Toxicol. **10**, 257-264.

17. Cooper, R. L., Stoker, T. E., Tyrey, L., Goldman, J. M., and McElroy, W. K. (2000). Atrazine disrupts the hypothalamic control of pituitary-ovarian function. Toxicol.Sci. **53**, 297-307.

18. Creange, J. E., Schane, H. P., Anzalone, A. J., and Potts, G. O. (1978). Interruption of pregnancy in rats by azastene, an inhibitor of ovarian and adrenal steroidogenesis. Fertil.Steril. **30**, 86-90.

19. de Araujo, L. F., Soares, J. M., Jr., Simoes, R. S., Calio, P. L., Oliveira-Filho, R. M., Simoes, M. J., Haidar, M. A., and Baracat, E. C. (2006). Effect of conjugated equine estrogens and tamoxifen administration on thyroid gland histomorphology of the rat. Clinics. **61**, 321-326.

20. DePaolo, L. V. (1985). Dose-dependent inhibition of the preovulatory surges of gonadotropins and prolactin by the antiestrogen CI-628: possible sites of action. Biol.Reprod. **32**, 745-752.

21. Dodds, E. C. and Lawson, W. (1936). Synthetic oestrogenic agents without the phenanthrene nucleus. Nature **137**, 996.

22. EDSTAC. 1998. Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) final report. U.S. Environmental Protection Agency.

23. Engelbregt, M. J., van Weissenbruch, M. M., Popp-Snijders, C., Lips, P., and Delemarre-van de Waal HA (2001). Body mass index, body composition, and

leptin at onset of puberty in male and female rats after intrauterine growth retardation and after early postnatal food restriction. Pediatr.Res. **50**, 474-478.

24. EVERETT, J. W. and SAWYER, C. H. (1950). A 24-hour periodicity in the "LH-release apparatus" of female rats, disclosed by barbiturate sedation. Endocrinology **47**, 198-218.

25. Feron, V. J., de Groot, A. P., Spanjers, M. T., and Til, H. P. (1973). An evaluation of the criterion "organ weight" under conditions of growth retardation. Food Cosmet.Toxicol. **11**, 85-94.

26. Goldman, J. M., Cooper, R. L., Rehnberg, G. L., Hein, J. F., McElroy, W. K., and Gray, L. E., Jr. (1986). Effects of low subchronic doses of methoxychlor on the rat hypothalamic-pituitary reproductive axis. Toxicol.Appl.Pharmacol. **86**, 474-483.

27. Goldman, J. M., Laws, S. C., Balchak, S. K., Cooper, R. L., and Kavlock, R. J. (2000). Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. Crit Rev.Toxicol. **30**, 135-196.

28. Gray, L. E., Jr., Ostby, J., Ferrell, J., Rehnberg, G., Linder, R., Cooper, R., Goldman, J., Slott, V., and Laskey, J. (1989). A dose-response analysis of methoxychlor-induced alterations of reproductive development and function in the rat. Fundam.Appl.Toxicol. **12**, 92-108.

29. Gupta, B. N., McConnell, E. E., Goldstein, J. A., Harris, M. W., and Moore, J. A. (1983). Effects of a polybrominated biphenyl mixture in the rat and mouse. I. Six-month exposure. Toxicol.Appl.Pharmacol. **68**, 1-18.

30. Hamilton, G. D. and Bronson, F. H. (1986). Food restriction and reproductive development: male and female mice and male rats. Am.J.Physiol **250**, R370-R376.

31. Hirsch, K. S., Adams, E. R., Hoffman, D. G., Markham, J. K., and Owen, N. V. (1986). Studies to elucidate the mechanism of fenarimol-induced infertility in male rats. Toxicol.Appl.Pharmacol. **86**, 391-399.

32. Hirsch, K. S., Jones, C. D., Lindstrom, T. D., Stamm, N. B., Sutton, G. P., Taylor, H. M., and Weaver, D. E. (1987a). Discovery and development of a novel class of nonsteroidal aromatase inhibitors. Steroids **50**, 201-217.

33. Hirsch, K. S., Weaver, D. E., Black, L. J., Falcone, J. F., and MacLusky, N. J. (1987b). Inhibition of central nervous system aromatase activity: a mechanism for fenarimol-induced infertility in the male rat. Toxicol.Appl.Pharmacol. **91**, 235-245.

34. Hodgson,E. (1987). Measurement of toxicity. In A textbook of modern toxicology (E.Hodgson and P.E.Levi, Eds.), pp. 252-253. Elsevier Science Publishing Co., Inc., Amsterdam, The Netherlands.

35. Kapoor, I. P., Metcalf, R. L., Nystrom, R. F., and Sangha, G. K. (1970). Comparative metabolism of methoxychlor, methiochlor, and DDT in mouse, insects, and in a model ecosystem. J.Agric.Food Chem. **18**, 1145-1152.

36. KENNEDY, G. C. and MITRA, J. (1963). Body weight and food intake as initiating factors for puberty in the rat. J.Physiol **166**, 408-418.

37. Kim, H. S., Kang, T. S., Kang, I. H., Kim, T. S., Moon, H. J., Kim, I. Y., Ki, H., Park, K. L., Lee, B. M., Yoo, S. D., and Han, S. Y. (2005). Validation study of OECD rodent uterotrophic assay for the assessment of estrogenic activity in Sprague-Dawley immature female rats. J.Toxicol.Environ.Health A **68**, 2249-2262.

38. Krishnan, A. V., Stathis, P., Permuth, S. F., Tokes, L., and Feldman, D. (1993). Bisphenol-A: an estrogenic substance is released from polycarbonate flasks during autoclaving. Endocrinology **132**, 2279-2286.

39. Laws, S. C. 9-26-2007. Results of perchlorate and thiram pubertal studies.

40. Laws, S. C., Carey, S. A., Ferrell, J. M., Bodman, G. J., and Cooper, R. L. (2000a). Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. Toxicol.Sci. **54**, 154-167.

41. Laws, S. C., Ferrell, J. M., Stoker, T. E., and Cooper, R. L. (2003). Pubertal development in female Wistar rats following exposure to propazine and atrazine biotransformation by-products, diamino-S-chlorotriazine and hydroxyatrazine. Toxicol.Sci. **76**, 190-200.

42. Laws, S. C., Ferrell, J. M., Stoker, T. E., Schmid, J., and Cooper, R. L. (2000b). The effects of atrazine on female wistar rats: an evaluation of the protocol for assessing pubertal development and thyroid function. Toxicol.Sci. **58**, 366-376.

43. Laws, S. C., Stoker, T. E., Ferrell, J. M., Hotchkiss, M. G., and Cooper, R. L. (2007). Effects of Altered Food Intake During Pubertal Development in Male and Female Wistar Rats. Toxicol.Sci.

44. Marinovich, M., Guizzetti, M., Ghilardi, F., Viviani, B., Corsini, E., and Galli, C. L. (1997). Thyroid peroxidase as toxicity target for dithiocarbamates. Arch.Toxicol. **71**, 508-512.

45. Marty, M. S., Crissman, J. W., and Carney, E. W. (1999). Evaluation of the EDSTAC female pubertal assay in CD rats using 17beta-estradiol, steroid biosynthesis inhibitors, and a thyroid inhibitor. Toxicol.Sci. **52**, 269-277.

46. Masutomi, N., Shibutani, M., Takagi, H., Uneyama, C., and Hirose, M. (2004). Dietary influence on the impact of ethinylestradiol-induced alterations in the endocrine/reproductive system with perinatal maternal exposure. Reprod.Toxicol. **18**, 23-33.

47. Masutomi, N., Shibutani, M., Takagi, H., Uneyama, C., Takahashi, N., and Hirose, M. (2003). Impact of dietary exposure to methoxychlor, genistein, or diisononyl phthalate during the perinatal period on the development of the rat endocrine/reproductive systems in later life. Toxicology **192**, 149-170.

48. McClain, R. M. (1989). The significance of hepatic microsomal enzyme induction and altered thyroid function in rats: implications for thyroid gland neoplasia. Toxicol.Pathol. **17**, 294-306.

49. McClain, R. M., Levin, A. A., Posch, R., and Downing, J. C. (1989). The effect of phenobarbital on the metabolism and excretion of thyroxine in rats. Toxicol.Appl.Pharmacol. **99**, 216-228.

50. McMullin, T. S., Andersen, M. E., Nagahara, A., Lund, T. D., Pak, T., Handa, R. J., and Hanneman, W. H. (2004). Evidence that atrazine and diaminochlorotriazine inhibit the estrogen/progesterone induced surge of luteinizing hormone in female sprague-dawley rats without changing estrogen receptor action. Toxicol.Sci. **79**, 278-286.

51. Merry, B. J. and Holehan, A. M. (1979). Onset of puberty and duration of fertility in rats fed a restricted diet. J.Reprod.Fertil. **57**, 253-259.

52. Merry, B. J. and Holehan, A. M. (1985). The endocrine response to dietary restriction in the rat. Basic Life Sci. **35**, 117-141.

53. Mycek, M. J. and Brezenoff, H. E. (1976). Tolerance to centrally administered phenobarbital. Biochem.Pharmacol. **25**, 501-504.

54. National Toxicology Program (NTP). 2007. Multigenerational reproductive toxicology study of ethinyl estradiol (CAS No 57-63-6) in Sprague-Dawley rats (Feed Studies). 547. Research Triangle Park, NC, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services. Technical Report Series.

55. NIEHS. 3-1-1997. Validation and regulatory acceptance of toxicological test methods: a report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (NIH 97-3981). DHHS/NTP/NIEHS/NTP.

56. O'Connor, J. C., Frame, S. R., Davis, L. G., and Cook, J. C. (1999). Detection of thyroid toxicants in a tier I screening battery and alterations in thyroid endpoints over 28 days of exposure. Toxicol.Sci. **51**, 54-70.

57. Odum, J., Lefevre, P. A., Tittensor, S., Paton, D., Routledge, E. J., Beresford, N. A., Sumpter, J. P., and Ashby, J. (1997). The rodent uterotrophic assay: critical protocol features, studies with nonyl phenols, and comparison with a yeast estrogenicity assay. Regul.Toxicol.Pharmacol. **25**, 176-188.

58. Odum, J., Tinwell, H., Jones, K., Van Miller, J. P., Joiner, R. L., Tobin, G., Kawasaki, H., Deghenghi, R., and Ashby, J. (2001). Effect of rodent diets on the sexual development of the rat. Toxicol.Sci. **61**, 115-127.

59. OECD. 1996. Final report of the OECD workshop on harmonization of validation and acceptance criteria for alternative toxicological test methods. Organisation for Economic Cooperation and Development.

60. OECD. 11-23-2006. OECD report of the validation of the rodent uterotrophic bioassay: Phase 2. Testing of potent and weak oestrogen agonists by multiple laboratories. 66. ENV/JM/MONO(2006)34. Paris, Organisation for Economic Co-operation and Development. Environment Directorate. OECD Environment, Health, and Safety Publications, Series on Testing and Assessment.

61. Ostby, J., Gray, L. E., Jr., Cooper, R. L., and Kelce, W. R. (1999). The estrogenic and antiandrogenic pesticide methoxychlor alters the reproductive tract and behavior without affecting pituitary size or LH and prolactin secretion in male rats. Toxicol.Ind.Health **15**, 37-47.

62. Pazos-Moura, C. C., Moura, E. G., Dorris, M. L., Rehnmark, S., Melendez, L., Silva, J. E., and Taurog, A. (1991). Effect of iodine deficiency and cold exposure on thyroxine 5'-deiodinase activity in various rat tissues. Am.J.Physiol **260**, E175-E182.

63. Perheentupa, A., Bergendahl, M., and Huhtaniemi, I. (1995). Modulation of gonadotropin secretion at the pituitary level by testosterone in gonadotropin-releasing hormone-treated male rats during food deprivation. Biol.Reprod. **52**, 808-813.

64. Reynolds, R. W. and Bryson, G. (1974). Effect of estradiol on the hypothalamic regulation of body weight in the rat. Res.Commun.Chem.Pathol.Pharmacol. **7**, 715-724.

65. Ronnekleiv, O. K., Ojeda, S. R., and McCann, S. M. (1978). Undernutrition, puberty and the development of estrogen positive feedback in the female rat. Biol.Reprod. **19**, 414-424.

66. Singh, M. M. and Kamboj, V. P. (1980). Estrogenic and antiestrogenic activities of two new nonsteroidal antifertility agents. Endokrinologie. **76**, 36-44.

67. Smith,A.G. (1991). Chlorinated hydrocarbon insecticides. In Handbook of pesticide toxicology (W.J.Hayes, Jr. and E.R.Laws, Jr., Eds.), Academic Press, Inc., New York, NY.

68. Stoker, T. E., Guidici, D. L., Laws, S. C., and Cooper, R. L. (2002). The effects of atrazine metabolites on puberty and thyroid function in the male Wistar rat. Toxicol.Sci. **67**, 198-206.

69. Stoker, T. E., Laws, S. C., Crofton, K. M., Hedge, J. M., Ferrell, J. M., and Cooper, R. L. (2004). Assessment of DE-71, a commercial polybrominated diphenyl ether (PBDE) mixture, in the EDSP male and female pubertal protocols. Toxicol.Sci. **78**, 144-155.

70. Stoker, T. E., Laws, S. C., Guidici, D. L., and Cooper, R. L. (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.

71. Thigpen, J. E., Haseman, J. K., Saunders, H. E., Setchell, K. D., Grant, M. G., and Forsythe, D. B. (2003). Dietary phytoestrogens accelerate the time of vaginal opening in immature CD-1 mice. Comp Med. **53**, 607-615.

72. Thigpen, J. E., Setchell, K. D., Saunders, H. E., Haseman, J. K., Grant, M. G., and Forsythe, D. B. (2004). Selecting the appropriate rodent diet for endocrine disruptor research and testing studies. ILAR.J. **45**, 401-416.

73. TULLNER, W. W. (1961). Uterotrophic action of the insecticide methoxychlor. Science **133**, 647.

74. Tyl, R. W., Myers, C. B., Marr, M. C., Thomas, B. F., Keimowitz, A. R., Brine, D. R., Veselica, M. M., Fail, P. A., Chang, T. Y., Seely, J. C., Joiner, R. L., Butala, J. H., Dimond, S. S., Cagen, S. Z., Shiotsuka, R. N., Stropp, G. D., and Waechter, J. M. (2002). Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. Toxicol.Sci. **68**, 121-146.

75. USEPA. 12-28-1998. Endocrine Disruptor Screening Program; Proposed statement of policy. Federal Register 63(248), 71542-71568. U.S. Environmental Protection Agency.

76. USEPA. 1999. Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. USEPA/SAB.

77. USEPA. 2007. Validation of screening and testing assays proposed for the EDSP. Version 5.4. USEPA/OPPTS/OSCP/EDSP.

78. Vinggaard, A. M., Jacobsen, H., Metzdorff, S. B., Andersen, H. R., and Nellemann, C. (2005). Antiandrogenic effects in short-term in vivo studies of the fungicide fenarimol. Toxicology **207**, 21-34.

79. Welch, R. M., Levin, W., and Conney, A. H. (1969). Estrogenic action of DDT and its analogs. Toxicol.Appl.Pharmacol. **14**, 358-367.

80. Widdowson, E. M. and McCance, R. A. (1960). Some effects of accelerating growth. I. General somatic development. Proc.R.Soc.Lond B Biol.Sci. **152**, 188-206.

81. Wolff, J. (1998). Perchlorate and the thyroid gland. Pharmacol.Rev. **50**, 89-105.

82. Yamasaki, K., Sawaki, M., Noda, S., Imatanaka, N., and Takatsuki, M. (2002a). Subacute oral toxicity study of ethynylestradiol and bisphenol A, based on the draft protocol for the "Enhanced OECD Test Guideline no. 407". Arch.Toxicol. **76**, 65-74.

83. Yamasaki, K., Sawaki, M., Noda, S., Wada, T., Hara, T., and Takatsuki, M. (2002b). Immature uterotrophic assay of estrogenic compounds in rats given diets of different phytoestrogen content and the ovarian changes with ICI 182,780 or antide. Arch.Toxicol. **76**, 613-620.

84. Yamasaki, K., Tago, Y., Nagai, K., Sawaki, M., Noda, S., and Takatsuki, M. (2002c). Comparison of toxicity studies based on the draft protocol for the 'Enhanced OECD Test Guideline no. 407' and the research protocol of 'Pubertal Development and Thyroid Function in Immature Male Rats' with 6-n-propyl-2-thiouracil. Arch.Toxicol. **76**, 495-501.

85. You, L., Casanova, M., Bartolucci, E. J., Fryczynski, M. W., Dorman, D. C., Everitt, J. I., Gaido, K. W., Ross, S. M., and Heck, H. H. (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.

86. Zoeller, R. T., Tan, S. W., and Tyl, R. W. (2007). General background on the hypothalamic-pituitary-thyroid (HPT) axis. Crit Rev.Toxicol. **37**, 11-53.

Appendix 1. Female pubertal protocol

Appendix 2. Detailed Review Paper

Appendix 3. Transferability study (TherImmune 1) summary report

Appendix 4. Transferability study (TherImmune 1) detailed table of results

Appendix 5. Multi-chemical study (RTI) summary report

Appendix 6. Multi-chemical study (RTI) detailed table of results

Appendix 7. Multi-chemical study (RTI) ANCOVA with body weight at weaning

Appendix 8. Multi-dose study (TherImmune 2) summary report

Appendix 9. Multi-dose study (TherImmune 2) detailed table of results

Appendix 10. Multi-dose study (TherImmune 2) ANCOVA with body weight at weaning

Appendix 11. White Paper on rat strain differences

Appendix 12. Reviewer's comments on White Paper on rat strain differences

Appendix 13. Interlaboratory validation study summary report (Charles River/Argus)

Appendix 14. Interlaboratory validation study summary report (Huntingdon)

Appendix 15. Interlaboratory validation study summary report (WIL)

Appendix 16. Interlaboratory validation study analysis report (Battelle)

Appendix 17. Interlaboratory validation study detailed table of results

Appendix 18. Interlaboratory validation study, comparison of results table